BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, C12N 15/00 — A1

(11) International Publication Number: WO 98/31837

(43) International Publication Date: 23 July 1998 (23.07.98)

(21) International Application Number: PCT/US98/00852

(22) International Filing Date: 16 January 1998 (16.01.98)

(30) Priority Data:
60/035,054 17 January 1997 (17.01.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/035,054 (CIP)
Filed on 17 January 1997 (17.01.97)

(71) Applicant *(for all designated States except US)*: MAXYGEN, INC. [US/US]; 3410 Central Expressway, Santa Clara, CA 95051 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DELCARDAYRE, Stephen, B. [US/US]; 101 Oak Rim Way #14, Los Gatos, CA 95032 (US). TOBIN, Mathew, B. [US/US]; 3450 Granada Avenue #76, Santa Clara, CA 95051 (US). STEMMER, Willem, P., C. [NL/US]; 108 Kathy Court, Los Gatos, CA 95030 (US). NESS, Jon, E. [US/US]; 1220 N. Fairoaks Avenue #3115, Sunnyvale, CA 94089 (US). MINSHULL, Jeremy [GB/US]; 11 Homer Lane, Menlo Park, CA 94025 (US). PATTEN, Phillip [US/US]; 2680 Fayette Drive #506, Mountain View, CA 94040 (US).

(74) Agents: LIEBESCHUETZ, Joe et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION

(57) Abstract

The invention provides methods employing iterative cycles of recombination and selection/screening for evolution of whole cells and organisms toward acquisition of desired properties. Examples of such properties include enhanced recombinogenicity, genome copy number, and capacity for expression and/or secretion of proteins and secondary metabolites.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION

CROSS-REFERENCE TO RELATED APPLICATION

This application derives priority from USSN 60/035,054, filed January 17, 1997, which is incorporated by reference in its entirety for all purposes.

TECHNICAL FIELD

The invention applies the technical field of molecular genetics to evolve the genomes of cells and organisms to acquire new and improved properties.

BACKGROUND

Cells have a number of well-established uses in molecular biology. For example, cells are commonly used as hosts for manipulating DNA in processes such as transformation and recombination. Cells are also used for expression of recombinant proteins encoded by DNA transformed into the cells. Some types of cells are also used as progenitors for generation of transgenic animals and plants. Although all of these processes are now routine, in general, the genomes of the cells used in these processes have evolved little from the genomes of natural cells, and particularly not toward acquisition of new or improved properties for use in the above processes.

The traditional approach to artificial or forced molecular evolution focuses on optimization of an individual gene having a discrete and selectable phenotype. The strategy is to clone a gene, identify a discrete function for the gene and an assay by which it can be selected, mutate selected positions in the gene (e.g., by error-prone PCR or cassette mutagenesis) and select variants of the gene for improvement in the known function of the gene. A variant having improved function can then be expressed in a desired cell type. This approach has a number of limitations. First, it is only

applicable to genes that have been isolated and functionally characterized. Second, the approach is usually only applicable to genes that have a discrete function. In other words, multiple genes that cooperatively confer a single phenotype cannot usually be optimized in this manner. Probably, most genes do have cooperative functions. Finally, this approach can only explore a very limited number of the total number of permutations even for a single gene. For example, varying even ten positions in a protein with every possible amino acid would generate 20^{10} variants, which is more than can be accommodated by existing methods of transfection and screening.

In view of these limitations, the traditional approach is inadequate for improving cellular genomes in many useful properties. For example, to improve a cell's capacity to express a recombinant protein might require modification in any or all of a substantial number of genes, known and unknown, having roles in transcription, translation, posttranslational modification, secretion or proteolytic degradation, among others. Attempting individually to optimize even all the known genes having such functions would be a virtually impossible task, let alone optimizing hitherto unknown genes which may contribute to expression in manners not yet understood.

The present invention provides *inter alia* novel methods for evolving the genome of whole cells and organisms which overcome the difficulties and limitations of prior methods.

DEFINITIONS

The term cognate refers to a gene sequence that is evolutionarily and functionally related between species. For example, in the human genome, the human CD4 gene is the cognate gene to the mouse CD4 gene, since the sequences and structures of these two genes indicate that they are highly homologous and both genes encode a protein which functions in signaling T cell activation through MHC class II-restricted antigen recognition.

Screening is, in general, a two-step process in which one first determines which cells do and do not express a screening marker and then physically separates the cells having the desired property. Selection is a form of screening in which identification and physical separation are achieved simultaneously by expression of a selection marker, which, in some genetic circumstances, allows cells expressing the marker to survive while other cells die (or vice versa). Screening markers include luciferase, β-galactosidase, and green fluorescent protein. Selection markers include drug and toxin resistance genes.

An exogenous DNA segment is one foreign (or heterologous) to the cell or homologous to the cell but in a position within the host cell nucleic acid in which the element is not ordinarily found. Exogenous DNA segments can be expressed to yield exogenous polypeptides.

The term gene is used broadly to refer to any segment of DNA associated with a biological function. Thus, genes include coding sequences and/or the regulatory sequences required for their expression. Genes also include nonexpressed DNA segments that, for example, form recognition sequences for other proteins.

Percentage sequence identity is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison. Optimal alignment of sequences for aligning a comparison window can be conducted by computerized implementations of algorithms GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, WI.

The term naturally-occurring is used to describe an object that can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified

by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

Asexual recombination is recombination occurring without the fusion of gametes to form a zygote.

SUMMARY OF THE CLAIMED INVENTION

In one aspect, the invention provides methods of evolving a cell to acquire a desired function. Such methods entail introducing a library of DNA fragments into a plurality of cells whereby at least one of the fragments undergoes recombination with a segment in the genome or an episome of the cells to produce modified cells. The modified cells are then screened for modified cells that have evolved toward acquisition of the desired function. DNA from the modified cells that have evolved toward the desired function is then recombined with a further library of DNA fragments at least one of which undergoes recombination with a segment in the genome or the episome of the modified cells to produce further modified cells. The further modified cells are then screened for further modified cells for further modified cells that have further evolved toward acquisition of the desired function. Steps of recombination and screening/selection are repeated as required until the further modified cells have acquired the desired function.

In some methods, the library or further library of DNA fragments is coated with recA protein to stimulate recombination with the segment of the genome. In some methods, the library of fragments is denatured to produce single-stranded DNA, the single-stranded DNA are annealed to produce duplexes some of which contain mismatches at points of variation in the fragments, and duplexes containing mismatches are selected by affinity chromatography to immobilized *Mut*S.

In some methods, the desired function is secretion of a protein, and the plurality of cells further comprises a construct encoding the protein. Optionally, the protein is toxic to the plurality of cells unless secreted, and the

modified or further modified cells having evolved toward acquisition of the desired function are screened by propagating the cells and recovering surviving cells.

In some methods, the desired function is enhanced recombination. In such methods, the library of fragments sometimes comprises a cluster of genes collectively conferring recombination capacity. Screening can be achieved using cells further comprises a gene encoding a marker whose expression is prevented by a mutation removable by recombination. The cells are screened by their expression of the marker resulting from removal of the mutation by recombination.

In some methods, the plurality of cells are plant cells and the desired property is improved resistance to a chemical or microbe, and in the screening the steps, the modified or further modified cells are exposed to the chemical or microbe and modified or further modified cells having evolved toward the acquisition of the desired function are selected by their capacity to survive the exposure.

In some methods, the plurality of cells are embryonic cells of an animal, and the method further comprises propagating the transformed cells to transgenic animals.

The invention further provides methods for performing *in vivo* recombination. These methods entail providing a cell incapable of expressing a cell septation gene. At least first and second segments from at least one gene are introduced into a cell, the segments differing from each other in at least two nucleotides, whereby the segments recombine to produce a library of chimeric genes. A chimeric gene is selected from the library having acquired a desired function.

The invention further provides methods of predicting efficacy of a drug in treating a viral infection. Such method entail recombining a nucleic acid segment from a virus, whose infection is inhibited by a drug, with at least a second nucleic acid segment from the virus, the second nucleic acid segment differing from the nucleic acid segment in at least two nucleotides, to produce a library of recombinant nucleic acid segments. Host cells are then contacted with a

collection of viruses having genomes including the recombinant nucleic acid segments in a media containing the drug, and progeny viruses resulting from infection of the host cells are collected.

A recombinant DNA segment from a first progeny virus recombines with at least a recombinant DNA segment from a second progeny virus to produce a further library of recombinant nucleic acid segments. Host cells are contacted with a collection of viruses having genomes including the further library or recombinant nucleic acid segments, in media containing the drug, and further progeny viruses are produced by the host cells. The recombination and selection steps are repeated, as necessary, until a further progeny virus has acquired a desired degree of resistance to the drug, whereby the degree of resistance acquired and the number of repetitions of needed to acquire it provide a measure of the efficacy of the drug in treating the virus.

The invention further provides methods of predicting efficacy of a drug in treating an infection by a pathogenic microorganism. These methods entail transforming a plurality of cells of the microorganism with a library of DNA fragments at least some of which undergo recombination with segments in the genome of the cells to produce modified microorganism cells. Modified microorganisms are propagated in a media containing the drug, and surviving microorganisms are recovered. DNA
from surviving microorganisms is recombined with a further library of DNA fragments at least some of which undergo recombination with cognate segments in the DNA from the surviving microorganisms to produce further modified microorganisms cells. Further modified microorganisms are propagated in media containing the drug, and further surviving microorganisms are collected. The recombination and selection steps are repeated as needed until a further surviving microorganism has acquired a desired degree of resistance to the drug, whereby the degree of resistance acquired and the number of repetitions of needed to acquire it provide a measure of the efficacy of the drug in killing the pathogenic

microorganism.

The invention further provides methods of evolving a cell to acquire a desired function. These methods entail providing a populating of different cells. The cells are cultured under conditions whereby DNA is exchanged between cells, forming cells with hybrid genomes. The cells are then screened or selected for cells that have evolved toward acquisition of a desired property. The DNA exchange and screening/selecting steps are repeated, as needed, with the screened/selected cells from one cycle forming the population of different cells in the next cycle, until a cell has acquired the desired property.

Mechanisms of DNA exchange include conjugation, phage-mediated transduction, protoplast fusion, and sexual recombination of the cells. Optionally, a library of DNA fragments can be transformed into the cells.

As noted, some methods of evolving a cell to acquire a desired property are effected by protoplast-mediated exchange of DNA between cells. Such methods entail forming protoplasts of a population of different cells. The protoplasts are then fused to form hybrid protoplasts, in which genomes from the protoplasts recombine to form hybrid genomes. The hybrid protoplasts are incubated under conditions promoting regeneration of cells. The next step is to select or screen to isolated regenerated cells that have evolved toward acquisition of the desired property. DNA exchange and selection/screening steps are repeated, as needed, with regenerated cells in one cycle being used to form protoplasts in the next cycle until the regenerated cells have acquired the desired property. Fungi are a preferred organism for conducting the above methods. Some methods further comprise a step of selecting or screening for fused protoplasts free from unfused protoplasts of parental cells. Some methods further comprise a step of selecting or screening for fused protoplasts with hybrid genomes free from cells with parental genomes. In some methods, protoplasts are provided by treating mycelia or spores with an enzyme that degrades cell walls. In some methods, the fungus is a fragile strain,

lacking capacity for intact cell wall synthesis, and protoplasts form spontaneously. In some methods, protoplasts are formed by treating mycelia with an inhibitor of cell wall formation to generate protoplasts.

In some methods, the desired property is expression and/or secretion of a protein or secondary metabolite, such as taxol. In some other methods, the desired property is capacity for meiosis. In some methods, the desired property is compatibility to form a heterokaryon with another strain.

The invention further provides methods of evolving a cell toward acquisition of a desired property. These methods entail providing a population of different cells. DNA is isolated from a first subpopulation of the different cells and encapsulated the in liposomes. Protoplasts are formed from a second subpopulation of the different cells. Liposomes are fused with the protoplasts, whereby DNA from the liposomes is taken up by the protoplasts and recombines with the genomes of the protoplasts. The protoplasts are incubated under regenerating conditions. Regenerating or regenerated cells are then selected or screened for evolution toward the desired property. The methods is then repeated with cells that have evolved toward the desired property in one cycle forming the population of different cells in the next cycle.

The invention further provides methods of evolving a cell toward acquisition of a desired property using artificial chromosomes. Such methods entail introducing a DNA fragment library cloned into an artificial chromosome into a population of cells. The cells are then cultured under conditions whereby sexual recombination occurs between the cells, and DNA fragments cloned into the artificial chromosome homologously recombines with corresponding segments of endogenous chromosomes of the populations of cells, and endogenous chromosomes recombine with each other. Cells that have evolved toward acquisition of the desired property are then selected or screened.

The invention further provides methods of evolving a DNA segment cloned into an artificial chromosome for acquisition of a desired property. These methods entail

providing a library of variants of the segment, each variant cloned into separate copies of an artificial chromosome. The copies of the artificial chromosome are introduced into a population of cells. The cells are cultured under conditions whereby sexual recombination occurs between cells and homologous recombination occurs between copies of the artificial chromosome bearing the variants. Variants are then screened or selected for evolution toward acquisition of the desired property.

The invention further provides hyperrecombinogenic recA proteins. Examples of such proteins clones 2, 4, 5, 6 and 13 shown in Fig. 13.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1: Scheme for *in vitro* shuffling of genes.

Fig. 2: Scheme for enriching for mismatched sequences using MutS.

Fig. 3: Alternative scheme for enriching for mismatched sequences using MutS.

Fig. 4: Scheme for evolving growth hormone genes to produce larger fish.

Fig. 5: Scheme for shuffling by protoplast fusion.

Fig. 6: Scheme for introducing a sexual cycle into fungi previously incapable of sexual reproduction.

Fig. 7. General scheme for shuffling of fung by protoplast fusion.

Fig. 8: Shuffling fungi by protoplast fusion with protoplasts generated by use of inhibitors of enzymes responsible for cell wall formation.

Fig. 9: Shuffling fungi by protoplast fusion using fungal strains deficient in cell-wall synthesis that spontaneously form protoplasts.

Fig. 10: YAC-mediated whole genome shuffling of Saccharomyces cerevisiae and related organisms.

Fig. 11: YAC-mediated shuffling of large DNA fragments.

Fig. 12: (A, B, C and D) DNA sequences of a wildtype recA protein (designated new Minshall) and five hyperrecombinogenic variants thereof.

Fig. 13: Amino acid sequences of a wildtype recA protein and five hyperrecombinogenic variants thereof.

DETAILED DESCRIPTION

I. General

A. The Basic Approach

The invention provides methods for artificially evolving cells to acquire a new or improved property by recursive sequence recombination. Briefly, recursive sequence recombination entails successive cycles of recombination and screening/selection to generate molecular diversity. That is, create a family of nucleic acid molecules showing substantial sequence and/or structural identity to each other but differing in the presence of mutations. Each recombination cycle is followed by at least one cycle of screening or selection for molecules having a desired characteristic. The molecule(s) selected in one round form the starting materials for generating diversity in the next round.

The cells to be evolved can be bacterial, archaebacteria, or eucaryotic and can constitute a homogeneous cell line or mixed culture. Suitable cells for evolution include the bacterial and eucaryotic cell lines commonly used in genetic engineering and protein expression. Suitable mammalian cells include those from, e.g., mouse, rat, hamster, primate, and human, both cell lines and primary cultures. Such cells include stem cells, including embryonic stem cells and hemopoietic stem cells, zygotes, fibroblasts, lymphocytes, Chinese hamster ovary (CHO), mouse fibroblasts (NIH3T3), kidney, liver, muscle, and skin cells. Other eucaryotic cells of interest include plant cells, such as maize, rice, wheat, cotton, soybean, sugarcane, tobacco, and arabidopsis; fish, algae, fungi (*penicillium, aspergillus, podospora, neurospora, saccharomyces*), insect (e.g., baculo lepidoptera), yeast (*picchia* and *saccharomyces, Schizosaccharomyces pombe*). Also of interest are many bacterial cell types, both gram-negative and gram-positive, such as *Bacillus subtilis, B. licehniformis, B. cereus, Escherichia coli, Pseudomonas, Salmonella, actinomycetes* and *Erwinia*. The complete genome

sequences of *E. coli* and *Bacillus subtilis* are described by Blattner et al., *Science* 277, 1454-1462 (1997); Kunst et al., *Nature* 390, 249-256 (1997)).

Evolution commences by generating a population of variant cells. Typically, the cells in the population are of the same type but represent variants of a progenitor cell. In some instances, the variation is natural as when different cells are obtained from different individuals within a species, or from different species. In other instances, variation is induced by mutagenesis of a progenitor cell. Mutagenesis can be effected by subjecting the cell to mutagenic agents, or if the cell is a mutator cell (e.g., has mutations in genes involved in DNA replication, recombination and/or repair which favor introduction of mutations) simply by propagating the mutator cells. Mutator cells can be generated from successive selections for simple phenotypic changes (e.g., acquisition of rifampicin-resistance, then nalidixic acid resistance then lac- to lac+ (see Mao et al., *J. Bacteriol.* 179, 417-422 (1997)).

In other instances, variation is the result of transferring into the cells (e.g., by conjugation, transformation, transduction or natural competence) a library of DNA fragments. At least one, and usually many of the fragments in the library, show some, but not complete, sequence or structural identity with a cognate or allelic gene within the cells sufficient to allow homologous recombination to occur. The library of fragments can derive from one or more sources. One source of fragments is a genomic library of fragments from a different species, cell type, organism or individual from the cells being transfected. In this situation, many of the fragments in the library have a cognate or allelic gene in the cells being transformed but differ from that gene due to the presence of naturally occurring species variation, polymorphisms, and mutations. Alternatively, the library can be derived from DNA from the same cell type as is being transformed after that DNA has been subject to induced mutation, by conventional methods, such as radiation, error-prone PCR, growth in a mutator organism or cassette

mutagenesis. In either of these situations, the genomic library can be a complete genomic library or subgenomic library deriving, for example, from a selected chromosome, or part of a chromosome or an episomal element within a cell. As well as, or instead of these sources of DNA fragments, the library can contain fragments representing natural or selected variants of selected genes of known function (i.e., focused libraries).

The number of fragments in a library can vary from a single fragment to about 10^{10}, with libraries having from 10^3 to 10^8 fragments being common. The fragments should be sufficiently long that they can undergo homologous recombination and sufficiently short that they can be introduced into a cell, and if necessary, manipulated before introduction. Fragment sizes can range from 10 b to 1000 kb, with sizes of 500-10,000 bases being common. Fragments can be double- or single-stranded.

The fragments can be introduced into cells as whole genomes or as components of viruses, plasmids, YACS, HACs or BACs or can be introduced as they are, in which case all or most of the fragments lack an origin of replication. Use of viral fragments with single-stranded genomes offer the advantage of delivering fragments in single stranded form, which promotes recombination. The fragments can also be joined to a selective marker before introduction. Inclusion of fragments in a vector having an origin of replication affords a longer period of time after introduction into the cell in which fragments can undergo recombination with a cognate gene before being degraded or selected against and lost from the cell, thereby increasing the proportion of cells with recombinant genomes. Optionally, the vector is a suicide vector capable of a longer existence than an isolated DNA fragment but not capable of permanent retention in the cell line. Such a vector can transiently express a marker for a sufficient time to screen for or select a cell bearing the vector, but is then degraded or otherwise rendered incapable of expressing the marker. The use of such vectors can be advantageous in performing subsequent rounds of recombination

to be discussed below. For example, some suicide vectors express a long-lived toxin which is neutralized by a short-lived molecule expressed from the same vector. Expression of the toxin alone will not allow vector to be established. Jense & Gerdes, *Mol. Microbiol.*, 17, 205-210 (1995); Bernard et al., *Gene* 162, 159-160. Alternatively, a vector can be rendered suicidal by incorporation of a defective origin of replication (i.e., temperature-sensitive) or by omission of an origin of replication. Vectors can also be rendered suicidal by inclusion of negative selection markers, such as oraB in yeast or sacB in many bacteria. These genes become toxic only in the presence of specific compounds. Such vectors can be selected to have a wide range of stabilities.

After introduction into cells, the fragments can recombine with DNA present in the genome or episomes of the cells by homologous, nonhomologous or site-specific recombination. For present purposes, homologous recombination makes the most significant contribution to evolution of the cells because this form of recombination amplifies the existing diversity between the DNA of the cells being transfected and the DNA fragments. For example, if a DNA fragment being transfected differs from a cognate or allelic gene at two positions, there are four possible recombination products, and each of these recombination products can be formed in different cells in the transformed population. Thus, homologous recombination of the fragment doubles the initial diversity in this gene. When many fragments recombine with corresponding cognate or allelic genes, the diversity of recombination products with respect to starting products increases exponentially with the number of fragments. Recombination results in modified cells having modified genomes and/or episomes.

The variant cells, whether the result of natural variation, mutagenesis, or recombination are screened or selected to identify a subset of cells that have evolved toward acquisition of a new or improved property. The nature of the screen, of course, depends on the property and several

examples will be discussed below. Optionally, the screening is repeated before performing subsequent cycles of recombination. Stringency can be increased in repeated cycles of screening.

The subpopulation of cells surviving screening are subjected to a further round of recombination. In some instances, the further round of recombination is effected by propagating the cells under conditions allowing exchange of DNA between cells. For example, protoplasts can be formed from the cells, allowed to fuse, and cells with recombinant genomes propagated from the fused protoplasts. Alternatively, exchange of DNA can be promoted by propagation of cells in an electric fied. For cells having a conjugative transfer apparatus, exchange of DNA can be promoted simply by propagating the cells.

In other methods, the further round of recombination is performed by a split and pool approach. That is, the surviving cells are divided into two pools. DNA is isolated from one pool, and if necessary amplified, and then transformed into the other pool. Accordingly, DNA fragments from the first pool constitute a further library of fragments and recombine with cognate fragments in the second pool resulting in further diversity.

In other methods, some or all of the cells surviving screening are transfected with a fresh library of DNA fragments, which can be the same or different from the library used in the first round of recombination. In this situation, the genes in the fresh library undergo recombination with cognate genes in the surviving cells. If genes are introduced as components of a vector, compatibility of this vector with any vector used in a previous round of transfection should be considered. If the vector used in a previous round was a suicide vector, there is no problem of incompatibility. If, however, the vector used in a previous round was not a suicide vector, a vector having a different incompatibility origin should be used in the subsequent round. In all of these formats, further recombination generates additional diversity in the DNA component of the cells resulting in further

modified cells.

The further modified cells are subjected to another round of screening/selection according to the same principles as the first round. Screening/selection identifies a subpopulation of further modified cells that have further evolved toward acquisition of the property. This subpopulation of cells can be subjected to further rounds of recombination and screening according to the same principles, optionally with the stringency of screening being increased at each round. Eventually, cells are identified that have acquired the desired property.

B. Variations

1. Coating Fragments with *rec*A Protein

The frequency of homologous recombination between library fragments and cognate endogenous genes can be increased by coating the fragments with a recombinogenic protein before introduction into cells. See Pati et al., *Molecular Biology of Cancer* 1, 1 (1996); Sena & Zarling, *Nature Genetics* 3, 365 (1996); Revet et al., *J. Mol. Biol.* 232, 779-791 (1993); Kowalczkowski & Zarling *in Gene Targeting* (CRC 1995), Ch. 7. The recombinogenic protein promotes homologous pairing and/or strand exchange. The best characterized *rec*A protein is from *E. coli* and is available from Pharmacia (Piscataway, NJ) . In addition to the wild-type protein, a number of mutant *rec*A-like proteins have been identified (e.g., *rec*A803). Further, many organisms have *rec*A-like recombinases with strand-transfer activities (e.g., Ogawa et al., *Cold Spring Harbor Symposium on Quantitative Biology* 18, 567-576 (1993); Johnson & Symington, *Mol. Cell. Biol.* 15, 4843-4850 (1995); Fugisawa et al., *Nucl. Acids Res.* 13, 7473 (1985); Hsieh et al., *Cell* 44, 885 (1986); Hsieh et al., *J. Biol. Chem.* 264, 5089 (1989); Fishel et al., *Proc. Natl. Acad. Sci. USA* 85, 3683 (1988); Cassuto et al., *Mol. Gen. Genet.* 208, 10 (1987); Ganea et al., *Mol. Cell Biol.* 7, 3124 (1987); Moore et al., *J. Biol. Chem.* 19, 11108 (1990); Keene et al., *Nucl. Acids Res.* 12, 3057 (1984); Kimiec, *Cold Spring Harbor Symp.* 48, 675 (1984); Kimeic, *Cell* 44, 545

(1986); Kolodner et al., *Proc. Natl. Acad. Sci. USA* 84, 5560 (1987); Sugino et al., *Proc. Natl. Acad. Sci. USA* 85, 3683 (1985); Halbrook et al., *J. Biol. Chem.* 264, 21403 (1989); Eisen et al., *Proc. Natl. Acad. Sci. USA* 85, 7481 (1988); McCarthy et al., *Proc. Natl. Acad. Sci. USA* 85, 5854 (1988); Lowenhaupt et al., *J. Biol. Chem.* 264, 20568 (1989). Examples of such recombinase proteins include *rec*A, *rec*A803, *uvs*X, (Roca, A.I., *Crit. Rev. Biochem. Molec. Biol.* 25, 415 (1990)), *sep*1 (Kolodner et al., *Proc. Natl. Acad. Sci. (U.S.A.)* 84, 5560 (1987); Tishkoff et al., *Molec. Cell. Biol.* 11, 2593), *Ruv*C (Dunderdale et al., *Nature* 354, 506 (1991)), *DST*2, *KEM*1, *XRN*1 (Dykstra et al., *Molec. Cell. Biol.* 11, 2583 (1991)), *STPα/DST*1 (Clark et al., *Molec. Cell. Biol.* 11, 2576 (1991)), *HPP*-1 (Moore et al., *Proc. Natl. Acad. Sci.* (U.S.A.) 88, 9067 (1991)), other eukaryotic recombinases (Bishop et al., *Cell* 69, 439 (1992); Shinohara et al., *Cell* 69, 457.

*Rec*A protein forms a nucleoprotein filament when it coats a single-stranded DNA. In this nucleoprotein filament, one monomer of *rec*A protein is bound to about 3 nucleotides. This property of *rec*A to coat single-stranded DNA is essentially sequence independent, although particular sequences favor initial loading of *rec*A onto a polynucleotide (e.g., nucleation sequences). The nucleoprotein filament(s) can be formed on essentially any to be shuffled and can form complexes with both single-stranded and double-stranded DNA in procaryotic and eucaryotic cells.

Before contacting with *rec*A or other recombinase, fragments are often denatured, e.g., by heat-treatment. *Rec*A protein is then added at a concentration of about 1-10 μM. After incubation, the *rec*A-coated single-stranded DNA is introduced into recipient cells by conventional methods, such as chemical transformation or electroporation. The fragments undergo homologous recombination with cognate endogenous genes. Because of the increased frequency of recombination due to recombinase coating, the fragments need not be introduced as components of vectors.

Fragments are sometimes coated with other nucleic acid binding proteins that promote recombination, protect

nucleic acids from degradation, or target nucleic acids to the nucleus. Examples of such proteins includes Agrobacterium virE2 (Durrenberger et al., *Proc. Natl. Acad. Sci. USA* 86, 9154-9158 (1989)).

2. *MutS* selection

The *E. coli* mismatch repair protein *MutS* can be used in affinity chromatography to enrich for fragments of double-stranded DNA containing at least one base of mismatch. The *MutS* protein recognizes the bubble formed by the individual strands about the point of the mismatch. See, e.g., Hsu & Chang, WO 9320233. The strategy of affinity enriching for partially mismatched duplexes can be incorporated into the present methods to increase the diversity between an incoming library of fragments and corresponding cognate or allelic genes in recipient cells.

Fig. 2 shows one scheme in which *MutS* is used to increase diversity. The DNA substrates for enrichment are substantially similar to each other but differ at a few sites. For example, the DNA substrates can represent complete or partial genomes (e.g., a chromosome library) from different individuals with the differences being due to polymorphisms. The substrates can also represent induced mutants of a wildtype sequence. The DNA substrates are pooled, restriction digested, and denatured to produce fragments of single-stranded DNA. The single-stranded DNA is then allowed to reanneal. Some single-stranded fragments reanneal with a perfectly matched complementary strand to generate perfectly matched duplexes. Other single-stranded fragments anneal to generate mismatched duplexes. The mismatched duplexes are enriched from perfectly matched duplexes by *MutS* chromatography (e.g., with *MutS* immobilized to beads). The mismatched duplexes recovered by chromatography are introduced into recipient cells for recombination with cognate endogenous genes as described above. *MutS* affinity chromatography increases the proportion of fragments differing from each other and the cognate endogenous gene. Thus, recombination between the incoming fragments and endogenous genes results in

greater diversity.

Fig. 3 shows a second strategy for *MutS* enrichment. In this strategy, the substrates for *MutS* enrichment represent variants of a relatively short segment, for example, a gene or cluster of genes, in which most of the different variants differ at no more than a single nucleotide. The goal of *MutS* enrichment is to produce substrates for recombination that contain more variations from each than sequences occurring in nature. This is achieved by fragmenting the substrates at random to produce overlapping fragments. The fragments are denatured and reannealed as in the first strategy. Reannealing generates some mismatched duplexes which can be separated from perfectly matched duplexes by *MutS* affinity chromatography. As before, *MutS* chromatography enriches for duplexes bearing at least a single mismatch. The mismatched duplexes are then reassembled into longer fragments. This is accomplished by cycles of denaturation, reannealing, and chain extension of partially annealed duplexes (see Section V). After several such cycles, fragments of the same length as the original substrates are achieved, except that these fragments differ from each other at multiple sites. These fragments are then introduced into cells where they undergo recombination with cognate endogenous genes.

3. Positive Selection For Allelic Exchange

The invention further provides methods of enriching for cells bearing modified genes relative to the starting cells. This can be achieved by introducing a DNA fragment library in a suicide vector (i.e., lacking a functional replication origin in the recipient cell type) containing both positive and negative selection markers. Optionally, multiple fragment libraries from different sources (e.g., *B. subtilis*, *B. licheniformis* and *B. cereus*) can be cloned into different vectors bearing different selection markers. Suitable positive selection markers include *neo*^{R}, kanamycin^R, *hyg*, *his*D, *gpt*, *ble*, *tet*^{R}, *hprt*, ura3 and sacB. Suitable negative selection markers include *hsv*-tk, *hprt*, *gpt*, and cytosine deaminase. Another strategy for applying negative selection

is to include a wildtype *rpsL* gene (encoding ribosomal protein S12) in a vector for use in cells having a mutant *rpsL* gene conferring streptomycin resistance. The mutant form of *rpsL* is recessive in cells having wildtype *rpsL*. Thus, selection for Sm resistance selects against cells having a wildtype copy of *rpsL*. See Skorupski & Taylor, *Gene* 169, 47-52 (1996). Alternatively, vectors bearing only a positive selection marker can be used with one round of selection for cells expressing the marker, and a subsequent round of screening for cells that have lost the marker (e.g., screening for drug sensitivity). The screen for cells that have lost the positive selection marker is equivalent to screening against expression of a negative selection marker. For example, Bacillus can be transformed with a vector bearing a CAT gene and a sequence to be integrated. See Harwood & Cutting, *Molecular Biological Methods for Bacillus*, at pp. 31-33. Selection for chloramphenicol resistance isolates cells that have taken up vector. After a suitable period to allow recombination, selection for CAT sensititivity isolates cells which have lost the CAT gene. About 50% of such cells will have undergone recombination with the sequence to be integrated.

Suicide vectors bearing a positive selection marker and optionally, a negative selection marker and a DNA fragment can integrate into host chromosomal DNA by a single crossover at a site in chromosomal DNA homologous to the fragment. Recombination generates an integrated vector flanked by direct repeats of the homologous sequence. In some cells, subsequent recombination between the repeats results in excision of the vector and either acquisition of a desired mutation from the vector by the genome or restoration of the genome to wildtype.

In the present methods, after transfer of the gene library cloned in a suitable vector, positive selection is applied for expression of the positive selection marker. Because nonintegrated copies of the suicide vector are rapidly eliminated from cells, this selection enriches for cells that have integrated the vector into the host chromosome. The cells surviving positive selection can then be propagated and

subjected to negative selection, or screened for loss of the positive selection marker. Negative selection selects against cells expressing the negative selection marker. Thus, cells that have retained the integrated vector express the negative marker and are selectively eliminated. The cells surviving both rounds of selection are those that initially integrated and then eliminated the vector. These cells are enriched for cells having genes modified by homologous recombination with the vector.

4. Individualized Optimization of Genes

In general, the above methods do not require knowledge of the number of genes to be optimized, their map location or their function. However, in some instances, where this information is available for one or more gene, it can be exploited. For example, if the property to be acquired by evolution is enhanced recombination of cells, one gene likely to be important is *recA*, even though many other genes, known and unknown, may make additional contributions. In this situation, the *recA* gene can be evolved, at least in part, separately from other candidate genes. The *recA* gene can be evolved by any of the methods of recursive recombination described in Section V. Briefly, this approach entails obtaining diverse forms of a *recA* gene, allowing the forms to recombine, selecting recombinants having improved properties, and subjecting the recombinants to further cycles of recombination and selection. At any point in the individualized improvement of *recA*, the diverse forms of *recA* can be pooled with fragments encoding other genes in a library to be used in the general methods described above. In this way, the library is seeded to contain a higher proportion of variants in a gene known to be important to the property sought to be acquired than would otherwise be the case.

5. Harvesting DNA Substrates for Shuffling

In some shuffling methods, DNA substrates are isolated from natural sources and are not easily manipulated by DNA modifying or polymerizing enzymes due to recalcitrant

impurities, which poison enzymatic reactions. Such difficulties can be avoided by processing DNA substrates through a harvesting strain. The harvesting strain is typically a cell type with natural competence and a capacity for homologous recombination between sequences with substantial diversity (e.g., sequences exhibiting only 75% sequence identity). The harvesting strain bears a vector encoding a negative selection marker flanked by two segments respectively complementary to two segments flanking a gene or other region of interest in the DNA from a target organism. The harvesting strain is contacted with fragments of DNA from the target organism. Fragments are taken up by natural competence, and a fragment of interest from the target organism recombines with the vector of the harvesting strain causing loss of the negative selection marker. Selection against the negative marker allows isolation of cells that have taken up the fragment of interest. Shuffling can be carried out in the harvester strain or vector can be isolated from the harvester strain for in vitro shuffling or transfer to a different cell type for in vivo shuffling. Alternatively, the vector can be transferred to a different cell type by conjugation, protoplast fusion or electrofusion. An example of a suitable harvester strain is *Acinetobacter calcoaceticus mut*S. Young et al., 97th ASM Meeting Abstracts. This strain is naturally competent and takes up DNA in a nonsequence-specific manner. Also, because of the mutS mutation, this strain is capable of homologous recombinatin of sequences showing only 75% sequence identity.

III. Applications

A. Recombinogenicity

One goal of whole cell evolution is to generate cells having improved capacity for recombination. Such cells are useful for a variety of purposes in molecular genetics including the *in vivo* formats of recursive sequence recombination described in Section V. Almost thirty genes (e.g., *rec*A, *rec*B, *rec*C, *rec*D, *rec*E, *rec*F, *rec*G, *rec*O, *rec*Q, *rec*R, *rec*T, *ruv*A, *ruv*B, *ruv*C, *sbc*B, *ssb*, *top*A, *gyr*A and B,

lig, *pol*A, *uvr*D, E, *rec*L, *mut*D, *mut*H, *mut*L, *mut*U,, *hel*D) and DNA sites (e.g., *chi*, *rec*N, *sbc*C) involved in genetic recombination have been identified in *E. coli*, and cognate forms of several of these genes have been found in other organisms (e.g., rad51, rad55 rad57, Dmc1 in yeast (see Kowalczykowski et al., *Microbiol. Rev.* 58, 401-465 (1994); Kowalczkowski & Zarling, *supra*) and human homologs of Rad51 and Dmc1 have been identified (see Sandler et al., *Nucl. Acids Res.* 24, 2125-2132 (1996)). At least some of the *E. coli* genes, including recA are functional in mammalian cells, and can be targeted to the nucleus as a fusion with SV40 large T antigen nuclear targeting sequence (Reiss et al., *Proc. Natl. Acad. Sci. USA*, 93, 3094-3098 (1996)). Further, mutations in mismatch repair genes, such as *mutL*, *mutS*, *mutH* relax homology requirements and allow recombination between more diverged sequences (Rayssiguier et al., *Nature* 342, 396-401 (1989)). The extent of recombination between divergent strains can be enhanced by impairing mismatch repair genes and stimulating SOS genes. Such can be achieved by use of appropriate mutant strains and/or growth under conditions of metabolic stress, which have been found to stimulate SOS and inhibit mismatch repair genes. Vulic et al., *Proc. Natl. Acad. Sci. USA* 94 (1997).

Starting substrates for recombination are selected according to the general principles described above. That is, the substrates can be whole genomes or fractions thereof containing recombination genes or sites. Large libraries of essentially random fragments can be seeded with collections of fragments constituting variants of one or more known recombination genes, such as recA. Alternatively, libraries can be formed by mixing variant forms of the various known recombination genes and sites.

The library of fragments is introduced into the recipient cells to be improved and recombination occurs, generating modified cells. The recipient cells preferably contain a marker gene whose expression has been disabled in a manner that can be corrected by recombination. For example, the cells can contain two copies of a marker gene bearing

mutations at different sites, which copies can recombine to generate the wildtype gene. A suitable marker gene is green fluorescent protein. A vector can be constructed encoding one copy of GFP having stopcodons near the N-terminus, and another copy of GFP having stopcodons near the C-terminus of the protein. The distance between the stop codons at the respective ends of the molecule is 500 bp and about 25% of recombination events result in active GFP. Expression of GFP in a cell signals that a cell is capable of homologous recombination to recombine in between the stop codons to generate a contiguous coding sequence. By screening for cells expressing GFP, one enriches for cells having the highest capacity for recombination. The same type of screen can be used following subsequent rounds of recombination. However, unless the selection marker used in previous round(s) was present on a suicide vector, subsequent round(s) should employ a second disabled screening marker within a second vector bearing a different origin of replication or a different positive selection marker to vectors used in the previous rounds.

B. Multigenomic Copy Number

The majority of bacterial cells in stationary phase cultures grown in rich media contain two, four or eight chromosomes. In minimal medium the cells contain one or two chromosomes. The number of chromosomes per bacterial cell thus depends on the growth rate of the cell as it enters stationary phase. This is because rapidly growing cells contain multiple replication forks, resulting in several chromosomes in the cells after termination. The number of chromosomes is strain dependent, although all strains tested have more than one chromosome in stationary phase. The number of chromosomes in stationary phase cells decreases with time. This appears to be due to fragmentation and degradation of entire chromosomes, similar to apoptosis in mammalian cells. This fragmentation of genomes in cells containing multiple genome copies results in massive recombination and mutagenesis. Useful mutants may find ways to use energy

sources that will allow them to continue growing.

Some cell types, such as *Deinococcus radians* (Daly and Minton *J. Bacteriol.* 177, 5495-5505 (1995)) exhibit polyploidy throughout the cell cycle. This cell type is highly radiation resistant due to the presence of many copies of the genome. High frequency recombination between the genomes allows rapid removal of mutations induced by a variety of DNA damaging agents.

A goal of the present methods is to evolve other cell types to have increased genome copy number akin to that of *Deinoccocus radians*. Preferably, the increased copy number is maintained through all or most of its cell cycle in all or most growth conditions. The presence of multiple genome copies in such cells results in a higher frequency of homologous recombination in these cells, both between copies of a gene in different genomes within the cell, and between a genome within the cell and a transfected fragment. The increased frequency of recombination allows the cells to be evolved more quickly to acquire other useful characteristics.

Starting substrates for recombination can be a diverse library of genes only a few of which are relevant to genomic copy number, a focused library formed from variants of gene(s) known or suspected to have a role in genomic copy number or a combination of the two. As a general rule one would expect increased copy number would be achieved by evolution of genes involved in replication and cell septation such that cell septation is inhibited without impairing replication. Genes involved in replication include *tus*, *xer*C, *xer*D, *dif*, *gyr*A, *gyr*B, *par*E, *par*C, *dif*, *Ter*A, *Ter*B, *Ter*C, *Ter*D, *Ter*E, *Ter*F, and genes influencing chromosome partitioning and gene copy number include *min*D, *muk*A (*tol*C), *muk*B, *muk*C, *muk*D, *spo*OJ, *spo*IIIE (Wake & Errington, *Annu. Rev. Genet.* 29, 41-67 (1995)). A useful source of substrates is the genome of a cell type such as *Deinoccocus radians* known to have the desired phenotype of multigenomic copy number. As well as or instead of the above substrates, fragments encoding protein or antisense RNA inhibitors to genes known to be involved in cell septation can also be used.

In nature, the existence of multiple genomic copies in a cell type would usually not be advantageous due to the greater nutritional requirements needed to maintain this copy number. However, artificial conditions can be devised to select for high copy number. Modified cells having recombinant genomes are grown in rich media (in which conditions, multicopy number should not be a disadvantage) and exposed to a mutagen, such as ultraviolet or gamma irradiation or a chemical mutagen, e.g., mitomycin, nitrous acid, photoactivated psoralens, alone or in combination, which induces DNA breaks amenable to repair by recombination. These conditions select for cells having multicopy number due to the greater efficiency with which mutations can be excised. Modified cells surviving exposure to mutagen are enriched for cells with multiple genome copies. If desired, selected cells can be individually analyzed for genome copy number (e.g., by quantitative hybridization with appropriate controls). Some or all of the collection of cells surviving selection provide the substrates for the next round of recombination. Eventually cells are evolved that have at least 2, 4, 6, 8 or 10 copies of the genome throughout the cell cycle.

C. Secretion

The protein (or metabolite) secretion pathways of bacterial and eukaryotic cells can be evolved to export desired molecules more efficiently, such as for the manufacturing of protein pharmaceuticals, small molecule drugs or specialty chemicals. Improvements in efficiency are particularly desirable for proteins requiring multisubunit assembly (such as antibodies) or extensive posttranslational modification before secretion.

The efficiency of secretion may depend on a number of genetic sequences including a signal peptide coding sequence, sequences encoding protein(s) that cleave or otherwise recognize the coding sequence, and the coding sequence of the protein being secreted. The latter may affect folding of the protein and the ease with which it can integrate into and traverse membranes. The bacterial

secretion pathway in *E. coli* include the *Sec*A, *Sec*B, *Sec*E, *Sec*D and *Sec*F genes. In *Bacillus subtilis*, the major genes are secA, secD, secE, secF, secY, ffh, ftsY together with five signal peptidase genes (sipS, sipT, sipU, sipV and sipW) (Kunst et al, *supra*). For proteins requiring posttranslational modification, evolution of genes effecting such modification may contribute to improved secretion. Likewise genes with expression products having a role in assembly of multisubunit proteins (e.g., chaperonins) may also contribute to improved secretion.

Selection of substrates for recombination follows the general principles discussed above. In this case, the focused libraries referred to above comprise variants of the known secretion genes. For evolution of procaryotic cells to express eucaryotic proteins, the initial substrates for recombination are often obtained at least in part from eucaryotic sources. Incoming fragments can undergo recombination both with chromosomal DNA in recipient cells and with the screening marker construct present in such cells (see below). The latter form of recombination is important for evolution of the signal coding sequence incorporated in the screening marker construct. Improved secretion can be screened by the inclusion of marker construct in the cells being evolved. The marker construct encodes a marker gene, operably linked to expression sequences, and usually operably linked to a signal peptide coding sequence. The marker gene is sometimes expressed as fusion protein with a recombinant protein of interest. This approach is useful when one wants to evolve the recombinant protein coding sequence together with secretion genes.

In one variation, the marker gene encodes a product that is toxic to the cell containing the construct unless the product is secreted. Suitable toxin proteins include diphtheria toxin and ricin toxin. Propagation of modified cells bearing such a construct selects for cells that have evolved to improve secretion of the toxin. Alternatively, the marker gene can encode a ligand to a known receptor, and cells bearing the ligand can be detected by FACS using labelled

receptor. Optionally, such a ligand can be operably linked to a phospholipid anchoring sequence that binds the ligand to the cell membrane surface following secretion. (See commonly owned, copending 08/309,345). In a further variation, secreted marker protein can be maintained in proximity with the cell secreting it by inoculating individual cells into agar drops. Secreted protein is confined within the agar matrix and can be detected by e.g., FACS^{tm}. In another variation, a protein of interest is expressed as a fusion protein together with b-lactamase or alkaline phosphatase. These enzymes metabolize commercially available chromogenic substrates (e.g., X-gal), but do so only after secretion into the periplasm. Appearance of colored substrate in a colony of cells therefore indicates capacity to secrete the fusion protein and the intensity of color is related to the efficiency of secretion.

The cells identified by these screening and selection methods have the capacity to secrete increased amounts of protein. This capacity may be attributable to increased secretion and increased expression, or from increased secretion alone.

D. Expression

Cells can also be evolved to acquire increased expression of a recombinant protein. The level of expression is, of course, highly dependent on the construct from which the recombinant protein is expressed and the regulatory sequences, such as the promoter, enhancer(s) and transcription termination site contained therein. Expression can also be affected by a large number of host genes having roles in transcription, posttranslational modification and translation. In addition, host genes involved in synthesis of ribonucleotide and amino acid monomers for transcription and translation may have indirect effects on efficiency of expression. Selection of substrates for recombination follows the general principles discussed above. In this case, focused libraries comprise variants of genes known to have roles in expression. For evolution of procaryotic cells to express

eucaryotic proteins, the initial substrates for recombination are often obtained, at least in part, from eucaryotic sources; that is eucaryotic genes encoding proteins such as chaperonins involved in secretion and/assembly of proteins. Incoming fragments can undergo recombination both with chromosomal DNA in recipient cells and with the screening marker construct present in such cells (see below).

Screening for improved expression can be effected by including a reporter construct in the cells being evolved. The reporter construct expresses (and usually secretes) a reporter protein, such as GFP, which is easily detected and nontoxic. The reporter protein can be expressed alone or together with a protein of interest as a fusion protein. If the reporter gene is secreted, the screening effectively selects for cells having either improved secretion or improved expression, or both.

E. Plant Cells

A further application of recursive sequence recombination is the evolution of plant cells, and transgenic plants derived from the same, to acquire resistance to pathogenic disease, chemicals, viricides, fungicides, insecticides (e.g., BT toxin), herbicides (e.g., atrazine or glyphosate) and bacteriocides. The substrates for recombination can again be whole genomic libraries, fractions thereof or focused libraries containing variants of gene(s) known or suspected to confer resistance to one of the above agents. Frequently, library fragments are obtained from a different kind of plant to the plant being evolved.

The DNA fragments are introduced into cultured plant cells or plant protoplasts by standard methods including electroporation (From et al., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., *Molecular Biology of Plant Tumors*, (Academic Press, New York, 1982) pp. 549-560; Howell, US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al.,

Nature 327, 70-73 (1987)), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* transformed with a Ti plasmid in which DNA fragments are cloned. The Ti plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and is stably integrated into the plant genome (Horsch et al., *Science* 233, 496-498 (1984); Fraley et al., *Proc. Natl. Acad. Sci. USA* 80, 4803 (1983)).

Diversity can also be generated by genetic exchange between plant protoplasts according to the same principles described below for fungal protoplasts. Procedures for formation and fusion of plant protoplasts are described by Takahashi et al., US 4,677,066; Akagi et al., US 5,360,725; Shimamoto et al., Us 5,250,433; Cheney et al., US 5,426,040.

After a suitable period of incubation to allow recombination to occur and for expression of recombinant genes, the plant cells are contacted with the agent to which resistance is to be acquired, and surviving plant cells are collected. Some or all of these plant cells can be subject to a further round of recombination and screening. Eventually, plant cells having the required degree of resistance are obtained.

These cells can then be cultured into transgenic plants. Plant regeneration from cultured protoplasts is described in Evans et al., "Protoplasts Isolation and Culture," *Handbook of Plant Cell Cultures* 1, 124-176 (MacMillan Publishing Co., New York, 1983); Davey, "Recent Developments in the Culture and Regeneration of Plant Protoplasts," *Protoplasts*, (1983) pp. 12-29, (Birkhauser, Basal 1983); Dale, "Protoplast Culture and Plant Regeneration of Cereals and Other Recalcitrant Crops," *Protoplasts* (1983) pp. 31-41, (Birkhauser, Basel 1983); Binding, "Regeneration of Plants," *Plant Protoplasts*, pp. 21-73, (CRC Press, Boca Raton, 1985).

In a variation of the above method, one or more preliminary rounds of recombination and screening can be performed in bacterial cells according to the same general strategy as described for plant cells. More rapid evolution can be achieved in bacterial cells due to their greater growth

rate and the greater efficiency with which DNA can be introduced into such cells. After one or more rounds of recombination/screening, a DNA fragment library is recovered from bacteria and transformed into the plants. The library can either be a complete library or a focused library. A focused library can be produced by amplification from primers specific for plant sequences, particularly plant sequences known or suspected to have a role in conferring resistance.

Example: Concatemeric Assembly of Atrazine-Catabolizing Plasmid *Pseudomonas* atrazine catabolizing genes AtzA and AtzB were subcloned from pMD1 (deSouza et al., *Appl. Environ. Microbiol.* 61, 3373-3378 (1995); de Souza et al., *J. Bacteriol.* 178, 4894-4900 (1996)) into pUC18. A 1.9 kb AvaI fragment containing AtzA was end-filled and inserted into an AvaI site of pUC18. A 3.9 kb ClaI fragment containing AtzB was end-filled and cloned into the HincII site of pUC18. AtzA was then excised from pUC18 with EcoRI and BamHI, AzB with BamHI and HindIII, and the two inserts were co-ligated into pUC18 digested with EcoRI and HindIII. The result was a 5.8 kb insert containing AtzA and AtzB in pUC18 (total plasmid size 8.4 kb).

Recursive sequence recombination was performed as follows. The entire 8.4 kb plasmid was treated with DNaseI in 50 mM Tris-Cl pH 7.5, 10 mM $MnCl_2$ and fragments between 500 and 2000 bp were gel purified. The fragments were assembled in a PCR reaction using Tth-XL enzyme and buffer from Perkin Elmer, 2.5 mM MgOAc, 400 μM dNTPs and serial dilutions of DNA fragments. The assembly reaction was performed in an MJ Research "DNA Engine" programmed with the following cycles:

1 94°C, 20 seconds
2 94°C, 15 seconds
3 40°C, 30 seconds
4 72°C, 30 seconds + 2 seconds per cycle
5 go to step 2, 39 more times
6 4°C

We were unable to amplify the AtzA and AtzB genes from the assembly reaction using the polymerase chain

reaction, so instead we purified DNA from the reaction by phenol extraction and ethanol precipitation, then digested the assembled DNA with a restriction enzyme that linearized the plasmid (KpnI: the KpnI site in pUC18 was lost during subcloning, leaving only the KpnI site in AtzA). Linearized plasmid was gel-purified, self-ligated overnight and transformed into *E. coli* strain NM522. (The choice of host strain was important: very little plasmid of poor quality was obtained from a number of other commercially available strains including TG1, DH10B, DH12S.)

Serial dilutions of the transformation reaction were plated onto LB plates containing 50 μg/ml ampicillin, the remainder of the transformation was made 25% in glycerol and frozen at -80oC. Once the transformed cells were titered, the frozen cells were plated at a density of between 200 and 500 on 150 mm diameter plates containing 500 μg/ml atrazine and grown at 37oC.

Atrazine at 500 μg/ml forms an insoluble precipitate. The products of the AtzA and AtzB genes transform atrazine into a soluble product. Cells containing the wild type AtzA and AtzB genes in pUC18 will thus be surrounded by a clear halo where the atrazine has been degraded. The more active the AtzA and AtzB enzymes, the more rapidly a clear halo will form and grow on atrazine-containing plates. Positives were picked as those colonies that most rapidly formed the largest clear zones. The (approximately) 40 best colonies were picked, pooled, grown in the presence of 50 μg/ml ampicillin and plasmid prepared from them. The entire process (from DNase-treatment to plating on atrazine plates) was repeated 4 times with 2000-4000 colonies/cycle.

A modification was made in the fourth round. Cells were plated on both 500 μg/ml atrazine, and 500 μg/ml of the atrazine analogue terbutylazine, which was undegradable by the wild type AtzA and AtzB genes. Positives were obtained that degraded both compounds. The atrazine chlorhydrolase (product of AtzA gene) was 10-100 fold higher than that produced by the wildtype gene).

F. Transgenic Animals

1. Transgene Optimization

One goal of transgenesis is to produce transgenic animals, such as mice, rabbits, sheep, pigs, goats, and cattle, secreting a recombinant protein in the milk. A transgene for this purpose typically comprises in operable linkage a promoter and an enhancer from a milk-protein gene (e.g., α, β, or γ casein, β-lactoglobulin, acid whey protein or α-lactalbumin), a signal sequence, a recombinant protein coding sequence and a transcription termination site. Optionally, a transgene can encode multiple chains of a multichain protein, such as an immunoglobulin, in which case, the two chains are usually individually operably linked to sets of regulatory sequences. Transgenes can be optimized for expression and secretion by recursive sequence recombination. Suitable substrates for recombination include regulatory sequences such as promoters and enhancers from milk-protein genes from different species or individual animals. Cycles of recombination can be performed *in vitro* or *in vivo* by any of the formats discussed in Section V. Screening is performed *in vivo* on cultures of mammary-gland derived cells, such as HC11 or MacT, transfected with transgenes and reporter constructs such as those discussed above. After several cycles of recombination and screening, transgenes resulting in the highest levels of expression and secretion are extracted from the mammary gland tissue culture cells and used to transfect embryonic cells, such as zygotes and embryonic stem cells, which are matured into transgenic animals.

2. Whole Animal Optimization

In this approach, libraries of incoming fragments are transformed into embryonic cells, such as ES cells or zygotes. The fragments can be variants of a gene known to confer a desired property, such as growth hormone. Alternatively, the fragments can be partial or complete genomic libraries including many genes.

Fragments are usually introduced into zygotes by microinjection as described by Gordon et al., *Methods Enzymol.*

101, 414 (1984); Hogan et al., *Manipulation of the Mouse Embryo: A Laboratory Manual* (C.S.H.L. N.Y., 1986) (mouse embryo); and Hammer et al., *Nature* 315, 680 (1985) (rabbit and porcine embryos); Gandolfi et al., *J. Reprod. Fert.* 81, 23-28 (1987); Rexroad et al., *J. Anim. Sci.* 66, 947-953 (1988) (ovine embryos) and Eyestone et al., *J. Reprod. Fert.* 85, 715-720 (1989); Camous et al., *J. Reprod. Fert.* 72, 779-785 (1984); and Heyman et al., *Theriogenology* 27, 5968 (1987) (bovine embryos) by reference in their entirety for all purposes). Zygotes are then matured, introduced and introduced into recipient female animals which gestate the embryo and give birth to a transgenic offspring.

Alternatively, transgenes can be introduced into embryonic stem cells (ES). These cells are obtained from preimplantation embryos cultured *in vitro*. Bradley et al., *Nature* 309, 255-258 (1984). Transgenes can be introduced into such cells by electroporation or microinjection. Transformed ES cells are combined with blastocysts from a nonhuman animal. The ES cells colonize the embryo and in some embryos form the germ line of the resulting chimeric animal. *See* Jaenisch, *Science*, 240, 1468-1474 (1988).

Regardless whether zygotes or ES are used, screening is performed on whole animals for a desired property, such as increased size and/or growth rate. DNA is extracted from animals having evolved toward acquisition of the desired property. This DNA is then used to transfect further embryonic cells. These cells can also be obtained from animals that have acquired toward the desired property in a split and pool approach. That is, DNA from one subset of such animals is transformed into embryonic cells prepared from another subset of the animals. Alternatively, the DNA from animals that have evolved toward acquisition of the desired property can be transfected into fresh embryonic cells. In either alternative, transfected cells are matured into transgenic animals, and the animals subjected to a further round of screening for the desired property.

Fig. 4 shows the application of this approach for evolving fish toward a larger size. Initially, a library is

prepared of variants of a growth hormone gene. The variants can be natural or induced. The library is coated with recA protein and transfected into fertilized fish eggs. The fish eggs then mature into fish of different sizes. The growth hormone gene fragment of genomic DNA from large fish is then amplified by PCR and used in the next round of recombination.

G. Rapid Evolution as a Predictive Tool

Recursive sequence recombination can be used to simulate natural evolution of pathogenic microorganisms in response to exposure to a drug under test. Using recursive sequence recombination, evolution proceeds at a faster rate than in natural evolution. One measure of the rate of evolution is the number of cycles of recombination and screening required until the microorganism acquires a defined level of resistance to the drug. The information from this analysis is of value in comparing the relative merits of different drugs and in particular, in predicting their long term efficacy on repeated administration.

The pathogenic microorganism used in this analysis include the bacteria that are a common source of human infections, such as *chlamydia, rickettsial bacteria, mycobacteria, staphylococci, treptocci, pneumonococci, meningococci* and *conococci, klebsiella, proteus, serratia, pseudomonas, legionella, diphtheria, salmonella, bacilli, cholera, tetanus, botulism, anthrax, plague, leptospirosis,* and *Lymes disease bacteria*. Evolution is effected by transforming an isolate of bacteria that is sensitive to a drug under test with a library of DNA fragments. The fragments can be a mutated version of the genome of the bacteria being evolved. If the target of the drug is a known protein, a focused library containing variants of the gene encoding that protein can be used. Alternatively, the library can come from other kinds of bacteria, especially bacteria typically found inhabiting human tissues, thereby simulating the source material available for recombination *in vivo*. The library can also come from bacteria known to be resistant to the drug. After transformation and propagation of bacteria

for an appropriate period to allow for recombination to occur and recombinant genes to be expressed, the bacteria are screened by exposing them to the drug under test and then collecting survivors. Surviving bacteria are subject to further rounds of recombination. The subsequent round can be effected by a split and pool approach in which DNA from one subset of surviving bacteria is introduced into a second subset of bacteria. Alternatively, a fresh library of DNA fragments can be introduced into surviving bacteria. Subsequent round(s) of selection can be performed at increasing concentrations of drug, thereby increasing the stringency of selection.

A similar strategy can be used to simulate viral acquisition of drug resistance. The object is to identify drugs for which resistance can be acquired only slowly, if at all. The viruses to be evolved are those that cause infections in humans for which at least modestly effective drugs are available. Substrates for recombination can come from induced mutants, natural variants of the same viral strain or different viruses. If the target of the drug is known (e.g., nucleotide analogs which inhibit the reverse transcriptase gene of HIV), focused libraries containing variants of the target gene can be produced. Recombination of a viral genome with a library of fragments is usually performed *in vitro*. However, in situations in which the library of fragments constitutes variants of viral genomes or fragments that can be encompassed in such genomes, recombination can also be performed *in vivo*, e.g., by transfecting cells with multiple substrate copies (see Section V). For screening, recombinant viral genomes are introduced into host cells susceptible to infection by the virus and the cells are exposed to a drug effective against the virus (initially at low concentration). The cells can be spun to remove any noninfected virus. After a period of infection, progeny viruses can be collected from the culture medium, the progeny viruses being enriched for viruses that have acquired at least partial resistance to the drug. Alternatively, virally infected cells can be plated in a soft agar lawn and

resistant viruses isolated from plaques. Plaque size provides some indication of the degree of viral resistance.

Progeny viruses surviving screening are subject to additional rounds of recombination and screening at increased stringency until a predetermined level of drug resistance has been acquired. The predetermined level of drug resistance may reflect the maximum dosage of a drug practical to administer to a patient without intolerable side effects. The analysis is particularly valuable for investigating acquisition of resistance to various combination of drugs, such as the growing list of approved anti-HIV drugs (e.g., AZT, ddI, ddC, d4T, TIBO 82150, nevaripine, 3TC, crixivan and ritonavir).

IV. Promotion of Genetic Exchange

(1) General

Some methods of the invention effect recombination of cellular DNA by propagating cells under conditions inducing exchange of DNA between cells. DNA exchange can be promoted by generally applicable methods such as electroporation, biolistics, cell fusion, or in some instances, by conjugation or agrobacterium mediated transfer. For example, *Agrobacterium* can transform *S. cerevisiae* with T-DNA, which is incorporated into the yeast genome by both homologous recombination and a gap repair mechanism. (Piers et al., *Proc. Natl. Acad. Sci. USA* 93(4), 1613-8 (1996)).

In some methods, initial diversity between cells (i.e., before genome exchange) is induced by chemical or radiation-induced mutagenesis of a progenitor cell type, optionally followed by screening for a desired phenotype. In other methods, diversity is natural as where cells are obtained from different individuals, strains or species.

In some shuffling methods, induced exchange of DNA is used as the sole means of effecting recombination in each cycle of recombination. In other methods, induced exchange is used in combination with natural sexual recombination of an organism. In other methods, induced exchange and/or natural sexual recombination are used in combination with the introduction of a fragment library. Such a fragment library

can be a whole genome, a whole chromosome, a group of functionally or genetically linked genes, a plasmid, a cosmid, a mitochondrial genome, a viral genome (replicative and nonreplicative) or fragments of any of these. The DNA can be linked to a vector or can be in free form. Some vectors contain sequences promoting homologous or nonhomologous recombination with the host genome. Some fragments contain double stranded breaks such as caused by shearing with glass beads, sonication, or chemical or enzymatic fragmentation, to stimulate recombination.

In each case, DNA can be exchanged between cells after which it can undergo recombination to form hybrid genomes. Cells bearing hybrid genomes are screened for a desired phenotype, and cells having this phenotype are isolated. These cells form the starting materials for the next cycle of recombination in a recursive recombination/selection scheme.

One means of promoting exchange of DNA between cells is by fusion of cells, such as by protoplast fusion. A protoplast results from the removal from a cell of its cell wall, leaving a membrane-bound cell that depends on an isotonic or hypertonic medium for maintaining its integrity. If the cell wall is partially removed, the resulting cell is strictly referred to as a spheroplast and if it is completely removed, as a protoplast. However, here the term protoplast includes spheroplasts unless otherwise indicated.

Protoplast fusion is described by Shaffner et al., *Proc. Natl. Acad. Sci. USA* 77, 2163 (1980) and other exemplary procedures are described by Yoakum et al., US 4,608,359, Takahashi et al., US 4,677,066 and Sambrooke et al., at Ch. 16. Protoplast fusion has been reported between strains, species, and genera (e.g., yeast and chicken erythrocyte).

Protoplasts can be prepared for both bacterial and eucaryotic cells, including mammalian cells and plant cells, by several means including chemical treatment to strip cell walls. For example, cell walls can be stripped by digestion with lysozyme in a 10-20% sucrose, 50 mM EDTA buffer. Conversion of cells to spherical protoplasts can be monitored

by phase-contrast microscopy. Protoplasts can also be prepared by propagation of cells in media supplemented with an inhibitor of cell wall synthesis, or use of mutant strains lacking capacity for cell wall formation. Preferably, eucaryotic cells are synchronized in G1 phase by arrest with inhibitors such as α-factor, *K. lactis* killer toxin, leflonamide and adenylate cyclase inhibitors. Optionally, some but not all, protoplasts to be fused can be killed and/or have their DNA fragmented by treatment with ultraviolet irradiation, hydroxylamine or cupferon (Reeves et al., *FEMS Microbiol. Lett.* 99, 193-198 (1992)). In this situation, killed protoplasts are referred to as donors, and viable protoplasts as acceptors. Using dead donors cells can be advantageous in subsequently recognizing fused cells with hybrid genomes, as described below. Further, breaking up DNA in donor cells is advantageous for stimulating recombination with acceptor DNA. Optionally, acceptor and/or fused cells can also be briefly, but nonlethally, exposed to uv irradiation further to stimulate recombination.

Once formed, protoplasts can be stabilized in a variety of osmolytes and compounds such as sodium chloride, potassium chloride, sodium phosphate, potassium phosphate, sucrose, sorbitol in the presence of DTT. The combination of buffer, pH, reducing agent, and osmotic stabilizer can be optimized for different cell types. Protoplasts can be induced to fuse by treatment with a chemical such as PEG, calcium chloride or calcium propionate or electrofusion (Tsoneva, *Acta Microbiologica Bulgaria* 24, 53-59 (1989)). A method of cell fusion employing electric fields has also been described. See Chang US, 4,970,154. Conditions can be optimized for different strains.

The fused cells are heterokaryons containing genomes from two component protoplasts. Fused cells can be enriched from unfused parental cells by sucrose gradient sedimentation or cell sorting. The two nuclei in the heterokaryons can fuse (karyogamy) and homologous recombination can occur between the genomes. The chromosomes can also segregate asymmetrically resulting in regenerated protoplasts that have lost or gained

whole chromosomes. The frequency of recombination can be increased by treatment with ultraviolet irradiation or by use of strains overexpressing recA or other recombination genes, such as MutS or MutL or the yeast rad genes, and cognate variants thereof in other species. Overexpression can be either the result of introduction of exogenous recombination genes or the result of selecting strains, which as a result of natural variation or induced mutation, overexpress endogenous recombination genes. The fused protoplasts are propagated under conditions allowing regeneration of cell walls, recombination and segregation of recombinant genomes into progeny cells from the heterokaryon and expression of recombinant genes. After, or occasionally before or during, recovery of fused cells, the cells are screened or selected for evolution toward a desired property.

Thereafter a subsequent round of recombination can be performed by preparing protoplasts from the cells surviving selection/screening in a previous round. The protoplasts are fused, recombination occurs in fused protoplasts, and cells are regenerated from the fused protoplasts. Protoplasts, regenerated or regenerating cells are subject to further selection or screening.

Alternatively, a subsequent round of recombination can be performed on a split pool basis as described above. That is, a first subpopulation of cells surviving selection/screening from a previous round are used for protoplast formation. A second subpopulation of cells surviving selection/screening from a previous round are used as a source for DNA library preparation. The DNA library from the second subpopulation of cells is then transformed into the protoplasts from the first subpopulation. The library undergoes recombination with the genomes of the protoplasts to form recombinant genomes. Cells are regenerated from protoplasts, and selection/screening is applied to regenerating or regenerated cells. In a further variation, a fresh library of nucleic acid fragments is introduced into protoplasts surviving selection/screening from a previous round.

An exemplary format for shuffling using protoplast fusion is shown in Fig. 5. The figure shows the following steps: protoplast formation of donor and recipient strains, heterokaryon formation, karyogamy, recombination, and segregation of recombination genomes into separate cells. Optionally, the recombinant genomes, if having a sexual cycle, can undergo further recombination with each other as a result of meiosis and mating. Cells are then screened or selected for a desired property. Cells surviving selection/screening are then used as the starting materials in a further cycle of protoplasting.

2. Selection For Hybrid Strains

The invention provides selection strategies to identify cells formed by fusion of components from parental cells from two or more distinct subpopulations. Selection for hybrid cells is usually performed before selecting or screening for cells that have evolved (as a result of genetic exchange) to acquisition of a desired property. A basic premise of most such selection schemes is that two initial subpopulations have two distinct markers. Cells with hybrid genomes can thus be identified by selection for both markers.

In one such scheme, at least one subpopulation of cells bears a selective marker attached to its cell membrane. Examples of suitable membrane markers include biotin, fluorescein and rhodamine. The markers can be linked to amide or thiol groups or through more specific derivatization chemistries, such as jodo-acetates, jodoacetamides, maleimides. For example, a marker can be attached as follows. Cells or protoplasts are washed with a buffer (e.g., PBS), which does not interfere with the chemical coupling of a chemically active ligand which reacts with amino groups of lysines or N-terminal aminogroups of membrane proteins. The ligand is either amine reactive itself (e.g., isothiocyanates, succinimidyl esters, sulfonyl chlorides) or is activated by a heterobifunctional linker (e.g. EMCS, SIAB, SPDP, SMB) to become amine reactive. The ligand is a molecule which is easily bound by protein derivatized magnetic beads or other

capturing solid supports. For example, the ligand can be succinimidyl activated biotin (Molecular probes: B-1606, B-2603, S-1515, S-1582). This linker is reacted with aminogroups of proteins residing in and on the surface of a cell. The cells are then washed to remove excess labelling agent before contacting with cells from the second subpopulation bearing a second selective marker.

The second subpopulation of cells can also bear a membrane marker, albeit a different membrane marker from the first subpopulation. Alternatively, the second subpopulation can bear a genetic marker. The genetic marker can confer a selective property such as drug resistance or a screenable property, such as expression of green fluorescent protein.

After fusion of first and second subpopulations of cells and recovery, cells are screened or selected for the presence of markers on both parental subpopulations. For example, fusants are enriched for one population by adsorbtion to psecific beads and these are then sorted by FACS^{tm} for those expressing a maker. Cells surviving both screen for both markers are those having undergone protoplast fusion, and are therefore more likely to have recombined genomes. Usually, the markers are screened or selected separately. Membrane-bound markers, such as biotin, can be screened by affinity enrichment for the cell membrane marker (e.g., by panning fused cells on an affinity matrix). For example, for a biotin membrane label, cells can be affinity purified using streptavidin-coated magnetic beads (Dynal). These beads are washed several times to remove the non-fused host cells. Alternatively, cells can be panned against an antibody to the membrane marker. In a further variation, if the membrane marker is fluorescent, cells bearing the marker can be identified by FACS^{tm}. Screens for genetic markers depend on the nature of the markers, and include capacity to grow on drug-treated media or FACS^{tm} selection for green fluorescent protein. If first and second cell populations have fluorescent markers of different wavelengths, both markers can be screened simultaneously by FACS^{tm} sorting.

In a further selection scheme for hybrid cells,

first and second populations of cells to be fused express different subunits of a heteromultimeric enzyme. Usually, the heteromultimeric enzyme has two different subunits, but heteromultimeric enzymes having three, four or more different subunits can be used. If an enzyme has more than two different subunits, each subunit can be expressed in a different subpopulation of cells (e.g., three subunits in three subpopulations), or more than one subunit can be expressed in the same subpopulation of cells (e.g., one subunit in one subpopulation, two subunits in a second subpopulation).

Hybrid cells representing a combination of genomes of first and second subpopulation component cells can then be recognized by an assay for intact enzyme. Such an assay can be a binding assay, but is more typically a functional assay (e.g., capacity to metabolize a substrate of the enzyme). Enzymatic activity can be detected for example by processing of a substrate to a product with a fluorescent or otherwise easily detectable emission spectrum. The individual subunits of a heteromultimeric enzyme used in such an assay preferably have no enzymic activity in dissociated form, or at least have significantly less activity in dissociated form than associated form. Preferably, the cells used for fusion lack an endogenous form of the heteromultimeric enzyme, or at least have significantly less endogenous activity than results from heteromultimeric enzyme formed by fusion of cells.

Penicillin acylase enzymes, cephalosporin acylase and penicillin acyltransferase are examples of suitable heteromultimeric enzymes. These enzymes are encoded by a single gene, which is translated as a proenzyme and cleaved by posttranslational autocatalytic proteolysis to remove a spacer endopeptide and generate two subunits, which associate to form the active heterodimeric enzyme. Neither subunit is active in the absence of the other subunit. However, activity can be reconstituted if these separated gene portions are expressed in the same cell by co-transformation. Other enzymes that can be used have subunits that are encoded by distinct genes (e.g., faoA amd faoB genes encode 3-oxoacyl-CoA thiolase of

Pseudonmonas fragi (*Biochem. J* 328, 815-820 (1997)).

An exemplary enzyme is penicillin G acylase from *Escherichia coli*, which has two subunits encoded by a single gene. Fragments of the gene encoding the two subunits operably linked to appropriate expression regulation sequences are transfected into first and second subpopulations of cells, which lack endogenous penicillin acylase activity. A cell formed by fusion of component cells from the first and second subpopulations expresses the two subunits, which assemble to form functional enzyme, e.g., penicillin acylase. Fused cells can then be selected on agar plates containing penicillin G, which is degraded by penicillin acylase.

In another variation, fused cells are identified by complementation of auxotrophic mutants. Parental subpopulations of cells can be selected with known auxotrophic mutations. Alternatively, auxotrophic mutations in a starting population of cells can be generated spontaneously by exposure to a mutagenic agent. Cells with auxotrophic mutations are selected by replica plating on minimal and complete media. Lesions resulting in auxotrophy are expected to be scattered throughout the genome, in genes for amino acid, nucleotide, and vitamin biosynthetic pathways. After fusion of parental cells, cells resulting from fusion can be identified by their capacity to grow on minimal media. These cells can then be screened or selected for evolution toward a desired property. Further steps of mutagenesis generating fresh auxotrophic mutations can be incorporated in subsequent cycles of recombination and screening/selection.

In variations of the above method, de novo generation of auxotrophic mutations in each round of shuffling can be avoided by resusing the same auxotrophs. For example, auxotrophs can be generated by transposon mutagensis using a transposon bearing selective marker. Auxotrophs are identified by a screen such as replica plating. Auxotrophs are pooled, and a generalized transducing phage lysate is prepared by growth of phage on a population of auxotrophic cells. A separate population of auxtrophic cells is subjected to genetic exchange, and complementation is used to selected

cells that have undergone genetic exchange and recombination. These cells are then screened or selected for acquisition of a desired property. Cells surviving screening or selection then have auxotrophic markers regenerated by introduction of the transducing transposon library. The newly generated auxotrophic cells can then be subject to further genetic exchange and screening/selection.

In a further variation, auxotrophic mutations are generated by homologous recombination with a targeting vector comprising a selective marker flanked by regions of homology with a biosynthetic region of the genome of cells to be evolved. Recombination between the vector and the genome inserts the positive selection marker into the genome causing an auxotrophic mutation. The vector is in linear form before introduction of cells. Optionally, the frequency of introduction of the vector can be increased by capping its ends with self-complementarity oligonucleotides annealed in a hair pin formation. Genetic exchange and screening/selection proceed as described above. In each round, targeting vectors are reintroduced regenerating the same population of auxotrophic markers.

In another variation, fused cells are identified by screening for a genomic marker present on one subpopulation of parental cells and an episomal marker present on a second subpopulation of cells. For example, a first subpopulation of yeast containing mitochondria can be used to complement a second subpopulation of yeast having a petite phenotype (i.e., lacking mitochondria).

In a further variation, genetic exchange is performed between two subpopulations of cells, one of which is dead. Viable cells are then screened for a marker present on the dead parental subpopulation.

3. Liposome-mediated transfers

In the methods noted above, in which nucleic acid fragment libraries are introduced into protoplasts, the nucleic acids are sometimes encapsulated in liposomes to facilitate uptake by protoplasts. Lipsome-mediated uptake of

DNA by protoplasts is described in Redford et al., *Mol. Gen. Genet.* 184, 567-569 (1981). Liposomes can efficiently deliver large volumes of DNA to protoplasts (see Deshayes et al., *EMBO J.* 4, 2731-2737 (1985)). Further, the DNA can be delivered as linear fragments, which are often more recombinogenic that whole genomes. In some methods, fragments are mutated prior to encapsulation in liposomes. In some methods, fragments are combined with RecA and homologs, or nucleases (e.g., restriction endonucleases) before encapsulation in liposomes to promote recombination.

4. Shuffling filamentous fungi

Filamentous fungi are particularly suited to performing the shuffling methods described above. Filamentous fungi are divided into four main classifications based on their structures for sexual reproduction: *Phycomycetes*, *Ascomycetes*, *Basidiomycetes* and the *Fungi Imperfecti*. *Phycomycetes* (e.g., *Rhizopus*, *Mucor*) form sexual spores in sporangium. The spores can be uni or multinucleate and often lack septated hyphae (coenocytic). *Ascomycetes* (e.g., *Aspergillus*, *Neurospora*, *Penicillum*) produce sexual spores in an ascus as a result of meiotic division. Asci typically contain 4 meiotic products, but some contain 8 as a result of additional mitotic division. *Basidiomycetes* include mushrooms, rusts and smuts and form sexual spores on the surface of a basidium. In *holobasidiomycetes*, such as mushrooms) the basidium is undivided. In *hemibasidiomycetes*, such as ruts (*Uredinales*) and smut fungi (*Ustilaginales*), the basidium is divided. *Fungi imperfecti*, include most human pathogens, have no sexual stage-vegetative reproduction.

Fungi can reproduce by asexual, sexual or parasexual means. Asexual reproduction, involves vegetative growth of mycelia, nuclear division and cell division without involvement of gametes and without nuclear fusion. Cell division can occur by sporulation, budding or fragmentation of hyphae.

Sexual reproduction provides a mechanism for shuffling genetic material between cells. A sexual

reproductive cycle is characterized by an alteration of a haploid phase and a diploid phase. Diploidy occurs when two haploid gamete nuclei fuse (karyogamy). The gamete nuclei can come from the same parental strains (self-fertile), such as in the homothallic fungi. In heterothallic fungi, the parental strains come from strains of different mating type.

A diploid cell converts to haploidy via meiosis, which essentially consists of two divisions of the nucleus accompanied by one division of the chromosomes. The products of one meiosis are a tetrad (4 haploid nuclei). In some cases, a mitotic division occurs after meiosis, giving rise to eight product cells. The arrangement of the resultant cells (usually enclosed in spores) resembles that of the parental strains. The length of the haploid and diploid stages differs in various fungi: for example, the Basidiomycetes and many of the *Ascomycetes* have a mostly hapolid life cycle (that is, meiosis occurs immediately after karyogamy), whereas others (e.g., *Saccharomyces cerevisiae*) are diploid for most of their life cycle (karyogamy occurs soon after meiosis). Sexual reproduction can occur between cells in the same strain (selfing) or between cells from different strains (outcrossing).

Sexual dimorphism (dioecism) is the separate production of male and female organs on different mycelia. This is a rare phenomenon among the fungi, although a few examples are known. Heterothallism (one locus-two alleles) allows for outcrossing between crosscompatable strains which are self-incompatable. The simplest form is the two allele-one locus system of mating types/factors, illustrated by the following organisms:

A and *a* in *Neurospora*

a and α in *Saccharomyces*

plus and minus in *Schizzosaccharomyces* and *Zygomycetes*

a_1 and a_2 in *Ustilago*

Multiple-allelomorph heterothallism is exhibited by some of the higher *Basidiomycetes* (e.g. *Gasteromycetes* and *Hymenomycetes*), which are heterothallic and have several mating types determined by multiple alleles. Heterothallism

in these organisms is either bipolar with one mating type factor, or tetrapolar with two unlinked factors, *A* and *B*. Stable, fertile heterokaryon formation depends on the presence of different A factors and, in the case of tetrapolar organisms, of different *B* factors as well. This system is effective in the promotion of outbreeding and the prevention of self-breeding. The number of different mating factors may be very large (i.e. thousands) (Kothe, *FEMS Microbiol. Rev.* 18, 65-87 (1996)), and non-parental mating factors may arise by recombination.

Parasexual reproduction provides a further means for shuffling genetic material between cells. This process allows recombination of parental DNA without involvement of mating types or games. Parasexual fusion occurs by hyphal fusion giving rise to a common cytoplasm containing different nuclei. The two nuclei can divided independently in the resulting heterokaryon but occasionally fuse. Fusion is followed by haploidization, which can involve loss of chromosomes and mitotic crossing over between homolgous chromosomes. Protoplast fusion is a form of parasexual reproduction.

Within the above four classes, fungi are also classified by vegetative compatibility group. Fungi within a vegetative compatibility group can form heterokaryons with each other. Thus, for exchange of genetic material between different strains of fungi, the fungi are usually prepared from the same vegetative compatibility group. However, some genetic exchange can occur between fungi from different incompatibility groups as a result of parasexual reproduction (see Timberlake et al., US 5,605,820). Further, as discussed elsewhere, the natural vegetative compatibility group of fungi can be expanded as a result of shuffling.

Several isolates of *Aspergillus nidulans*, *A. flavus*, *A. fumigatus*, *Penicillium chrysogenum*, *P. notatum*, *Cephalosporium chrysogenum*, *Neurospora crassa*, *Aureobasidium pullulans* have been karyotyped. Genome sizes generally range between 20 and 50 Mb among the Aspergilli. Differences in karyotypes often exist between similar strains and are also caused by transformation with exogenous DNA. Filamentous

fungal genes contain introns, usually ~50-100 bp in size, with similar consensus 5' and 3' splice sequences. Promotion and termination signals are often cross-recognizable, enabling the expression of a gene/pathway from one fungus (e.g. *A. nidulans*) in another (e.g. *P. chrysogenum*).

The major components of the fungal cell wall are chitin (or chitosan), β-glucan, and mannoproteins. Chitin and β-glucan form the scaffolding, mannoproteins are interstitial components which dictate the wall's porosity, antigenicity and adhesion. Chitin synthetase catalyzes the polymerization of β-(1,4)-linked N-acetylglucosamine (GIcNAc) residues, forming linear strands running antiparallel; β-(1,3)-glucan synthetases catalyze the homopolymerization of glucose.

One general goal of shuffling is to evolve fungi to become useful hosts for genetic engineering, in particular for the shuffling of unrelated genes. *A. nidulans* is generally the fungal organism of choice to serve as a host for such manipulations because of its sexual cycle and well-established use in classical and molecular genetics. Another general goal is to improve the capacity of fungi to make specific compounds (e.g. antibacterials (penicillins, cephalosporins), antifungals (e.g. echinocandins, aureobasidins), and wood-degrading enzymes). There is some overlap between these general goals, and thus, some desired properties are useful for achieving both goals.

One desired property is the introduction of meiotic apparati into fungi presently lacking a sexual cycle (see Sharon et al., *Mol. Gen. Genet.* 251, 60-68 (1996)). A scheme for introducing a sexual cycle into the fungi *P. chrysogenum* (a fungus imperfecti) is shown in Fig. 6. Subpopulations of protoplasts are formed from *A. nidulans* (which has a sexual cycle) and *P. chrysogenum*, which does not. The two strains preferably bear different markers. The *A. nidulans* protoplasts are killed by treatment with uv or hydroxylamine. The two subpopulations are fused to form heterokaryons. In some heterokaryons, nuclei fuse, and and some recombination occurs. Fused cells are cultured under conditions to generate new cell walls and then to allow sexual recombination to

occur. Cells with recombinant genomes are then selected (e.g., by selecting for complementation of auxotrophic markers present on the respective parent strains). Cells with hybrid genomes are more likely to have acquired the genes necessary for a sexual cycle. Protoplasts of cells can then be crossed with killed protoplasts of a further population of cells known to have a sexual cycle (the same or different as the previous round) in the same manner, followed by selection for cells with hybrid genomes.

Another desired property is the production of a mutator strain of fungi. Such a fungus can be produced by shuffling a fungal strain containing a marker gene with one or more mutations that impair or prevent expression of a functional product. Shufflants are propagated under conditions that select for expression of the positive marker (while allowing a small amount of residual growth without expression). Shufflants growing fastest are selected to form the starting materials for the next round of shuffling.

Another desired property is to expand the host range of a fungus so it can form heterokaryons with fungi from other vegetative compatibility groups. Incompatability between species results from the interactions of specific alleles at different incompatability loci (such as the "*het*" loci). If two strains undergo hyphal anastomosis, a lethal cytoplasmic incompatability reaction may occur if the strains differ at these loci. Strains must carry identical loci to be entirely compatible. Several of these loci have been identified in various species, and the incompatibility effect is somewhat additive (hence, "partial incompatibility" can occur). Some tolerant and *het*-negative mutants have been described for these organisms (e.g. Dales & Croft, *J. Gen. Microbiol.* 136, 1717-1724 (1990)). Further, a tolerance gene (tol) has been reported, which suppresses mating-type heterokaryon incompatibility. Shuffling is performed between protoplasts of strains from different incompatibility groups. A preferred format uses a live acceptor strain and a UV-irradiated dead acceptor strain. The UV irradiation serves to introduce mutations into DNA inactivating het genes. The two strains

should bear different genetic markers. Protoplasts of the strain are fused, cells are regenerated and screened for complementation of markers. Subsequent rounds of shuffling and selection can be performed in the same manner by fusing the cells suriving screening with a protoplasts of a fresh population of donor cells.

Another desired property is the introduction of multiple-allelomorph heterothallism into *Ascomycetes* and *Fungi imperfecti*, which do not normally exhibit this property. This mating system allows outbreeding without self-breeding. Such a mating system can be introduced by shuffling *Ascomycetes* and *Fungi imperfecti* with DNA from *Gasteromycetes* or *Hymenomycetes*, which have such a system.

Another desired property is spontaneous formation of protoplasts to facilitate use of a fungal strain as a shuffling host. Here, the fungus to be evolved is typically mutagenized. Spores of the fungus to be evolved are briefly treated with a cell-wall degrading agent for a time insufficient for complete protoplast formation, and are mixed with protoplasts from other strain(s) of fungi. Protoplasts formed by fusion of the two different subpopulations are identified by genetic or other selection/or screening as described above. These protoplasts are used to regenerate mycelia and then spores, which form the starting material for the next round of shuffling. In the next round, at least some of the surviving spores are treated with cell-wall removing enzyme but for a shorter time than the previous round. After treatment, the partially stripped cells are labelled with a first label. These cells are then mixed with protoplasts, which may derive from other cells surviving selection in a previous round, or from a fresh strain of fungi. These protoplasts are physically labelled with a second label. After incubating the cells under conditions for protoplast fusion fusants with both labels are selected. These fusants are used to generate mycelia and spores for the next round of shuffling, and so forth. Eventually, progeny that spontaneously form protoplasts (i.e., without addition of cell wall degrading agent) are identified.

Another desired property is the acquisition and/or improvement of genes encoding enzymes in biosynthetic pathways, genes encoding transporter proteins, and genes encoding proteins involved in metabolic flux control. In this situation, genes of the pathway can be introduced into the fungus to be evolved either by genetic exchange with another strain of fungus possessing the pathway or by introduction of a fragment library from an organism possessing the pathway. Genetic material of these fungi can then be subjected to further shuffling and screening/selection by the various procedures discussed in this application. Shufflant strains of fungi are selected/screened for production of the compound produced by the metabolic pathway or precursors thereof.

Another desired property is increasing the stability of fungi to extreme conditions such as heat. In this situation, genes conferring stability can be acquired by exchanging DNA with or transforming DNA from a strain that already has such properties. Alternatively, the strain to be evolved can be subjected to random mutagenesis. Genetic material of the fungus to be evolved can be shuffled by any of the procedures described in this application, with shufflants being selected by surviving exposure to extreme conditions.

Another desired property is capacity of a fungus to grow under altered nutritional requirements (e.g., growth on particular carbon or nitrogen sources). Altering nutritional requirements is particularly valuable, e.g., for natural isolates of fungi that produce valuable commercial products but have esoteric and therefore expensive nutritional requirement. The strain to be evolved undergoes genetic exchange and/or transformation with DNA from a strain that has the desired nutritional requirements. The fungus to be evolved can then optionally be subjected to further shuffling as described in this application and with recombinant strains being selected for capacity to grow in the desired nutritional circumstances. Optionally, the nutritional circumstances can be varied in successive rounds of shuffling starting at close to the natural requirements of the fungus to be evolved and in subsequent rounds approaching the desired nutritional

requirements.

Another desired property is acquisition of natural competence in a fungus. The procedure for acquisition of natural competence by shuffling is generally described in PCT/US97/04494. The fungus to be evolved typically undergoes genetic exchange or transformation with DNA from a bacterial strain or fungal strain that already has this property. Cells with recombinant genomes are then selected by capacity to take up a plasmid bearing a selective marker. Further rounds of recombination and selection can be performed using any of the procedures described above.

Another desired property is reduced or increased secretion of proteases and DNase. In this situation, the fungus to be evolved can acquire DNA by exchange or transformation from another strain known to have the desired property. Alternatively, the fungus to be evolved can be subject to random mutagenesis. The fungus to be evolved is shuffled as above. Before selection/screening isolates, pooled isolates of fungi are typically lysed to release proteases or DNase to the surrounding media. The presence of such enzymes, or lack thereof, can be assayed by contacting the media with a fluorescent molecule tethered to a support via a peptide or DNA linkage. Cleavage of the linkage releases detectable fluorescence to the media.

Another desired property is producing fungi with altered transporters (e.g., MDR). Such altered transporters are useful, for example, in fungi that have been evolved to produce new secondary metabolites, to allow entry of precursors required for synthesis of the new secondary metabolites into a cell, or to allow efflux of the secondary metabolite from the cell. Transporters can be evolved by introduction of a library of transporter variants into a fungal cells and allowing the cells to recombine by sexual or parasexual recombination. To evolve a transporter with capacity to transport a precursor into the cells, cells are propagated in the present of precursor, and cells are then screened for production of metabolite. To evolve a transporter with capacity to export a metabolite, cells are

propagated under conditions supporting production of the metabolite, and screened for export of metabolite to culture medium.

A general method of fungal shuffling is shown in Fig. 7. Spores from a frozen stock or fresh from an agar plate are used to inoculate suitable liquid medium (1). Spores are germinated resulting in hyphal growth (2). Mycelia are harvested, and washed by filtration and/or centrifugation. Optionally the sample is pretreated with DTT to enhance protoplast formation (3). Protoplasting is performed in an osmotically stabling medium (e.g., 1 m NaCl/20mM MgSO4, pH 5.8) by the addition of cell wall-degrading enzyme (e.g., Novozyme 234) (4). Cell wall degrading enzyme is removed by repeated washing with osmotically stabilizing solution (5). Protoplasts can be separated from mycelia, debris and spores by filtration through miracloth, and density centrifugation (6). Protoplasts are harvested by centrifugation and resuspended to the appropriate concentration. This step may lead to some protoplast fusion (7). Fusion can be stimulated by addition of PEG (e.g., PEG 3350), and/or repeated centrifugation and resuspension with or without PEG. Electrofusion can also be performed (8). Fused protoplasts can optionally be enriched from unfused protoplasts by sucrose gradient sedimentation (or other methods of screening described above). Fused protoplasts can optionally be treated with ultraviolet irradiation to stimulate recombination (9). Protoplasts are cultured on osmotically stabilized agar plates to regenerate cell walls and form mycelia (10). The mycelia are used to generate spores (11), which are used as the starting material in the next round of shuffling (12). Selection for a desired property can be performed either on regenerated mycelia or spores derived therefrom.

In an alternative method, protoplasts are formed by inhibition of one or more enzymes required for cell wall synthesis (see Fig. 8). The inhibitor should be fungistatic rather than fungicidal under the conditions of use. Examples of inhibitors include antifungal compounds described by (see Georgopapadakou & Walsh, *Antimicrob. Ag. Chemother.* 40,

279-291 (1996); Lyman & Walsh, *Drugs* 44, 9-35 (1992)). Other examples include chitin synthase inhibitors (polyoxin or nikkornycin compounds) and/or glucan synthase inhibitors (e.g. echinocandins, papulocandins, pneurnocandins). Inhibitors should be applied in osmotically stabilized medium. Cells stripped of their cell walls can be fused or otherwise employed as donors or hosts in genetic transformation/strain development programs. A possible scheme utilizing this method reiteratively is outlined in Figure 8.

In a further variation, protoplasts are prepared using strains fungi, which are genetically deficient or compromised in their ability to synthesize intact cell walls (see Fig. 9). Such mutants are generally referred to as fragile, osmotic-remedial, or cell wall-less, and can be obtainable from strain depositories. Examples of such strains include *Neurospora crassa os* mutants (Selitrennikoff, *Antimicrob. Agents. Chemother.* 23, 757-765 (1983)). Some such mutations are temperature-sensitive. Temperature-sensitive strains can be propagated at the permissive temperature for purposes of selection and amplification and at a nonpermissive temperature for purposes of protoplast formation and fusion.A temperature sensitive strain *Neurospora crassa* os strain has been described which propagates as protoplasts when growth in osmotically stabilizing medium containing sorbose and polyoxin at nonpermissive temperature but generates whole cells on transfer to medium containing sorbitol at a permissive temperature. See US 4,873,196.

Other suitable strains can be produced by targeted mutagenesis of genes involved chitin synthesis, glucan synthesis and other cell wall-related processes. Examples of such genes include CHT1, CHT2 and CALI (or CSD2) of *Saccharomyces cerevisiae* and *Candida spp.* (Georgopapadakou & Walsh 1996); *ETGI/FKSI/CNDI/ CWH53/PB RI* and homologs in *S. cerevisiae, Candida albicans, Cryptococcus neoformans, Aspergillus fumigatus, ChvAINdvA Agrobacterium* and *Rhizobium*. Other examples are *MA, orl*B, *orIC, MD, ts*E, and *bim*G of *Aspergillus nidulans* (Borgia, *J. Bacteriol.* 174, 377-389 (1992)). OrlA 1, tse6 and bimG11 mutant strains have

mutations resulting in lysis at restrictive temperatures. Lysis is prevented by osmotic stabilization. Mutation is complemented by addition of N-acetylglucosamine (GlcNAc). bimG11 mutant strains are ts for a type 1 protein phosphatase in conidia. Other suitable genes are *chs*A, *chs*B, *chs*C, *chs*D and *chs*E of *Aspergillus fumigatus*; *chs1 and chs2 of Neurospora crassa* ; *Phycomyces blakesleeanus MM* and chs1, 2 and 3 of *S. cerevisiae*. Chs1 is a non-essential repair enzyme; chs2 is involved in septum formation and chs3 is involved in cell wall maturation and bud ring formation. Other useful strains are S. cerevisiae CLY mutant strains (cell lysis) ts strains (Paravicini et al., *Mol. Cell Biol.* 12, 4896-4905 (1992)), such as a deletion of the PKC 1 gene (CLY 15 strain), a strain VY 1160-ts mutation in srb (actin gene) (Schade et al. *Acta Histochem. Suppl.* 41, 193-200 (1991)), and ses, haploid mutants with increased sensitivity to cell-wall digesting enzymes isolated from snail gut (Metha & Gregory, *Appl. Environ. Microbiol.* 41, 992-999 (1981)). Other useful strains are *C. albicans* chs 1, 2, 3 chitin synthetases, osmotic remedial conditional lethal mutants (Payton & de Tiani, *Curr. Genet.* 17, 293-296 (1990)); *C. utilis* mutants with increased sensitivity to cell-wall digesting enzymes isolated from snail gut (Metha & Gregory, 1981, *supra*); and *N. crassa* mutants os-1,2,3,4,5,6 (Selitrennikoff, *Antimicrob. Agents Chemother.* 23, 757-765 (1983)). Such mutants grow and divide without a cell wall at 37°C, but at 22°C produce a cell wall.

Targeted mutagenesis can be achieved by transforming cells with a positive-negative selection vector containing homologous regions flanking a segment to be targeted, a positive selection marker between the homologous regions and a negative selection marker outside the homologous regions (see Capecchi, US 5,627,059). In a variation, the negative selection marker can be an antisense transcript of the positive selection marker (see US 5,527,674).

Other suitable cells can be selected by random mutagenesis or shuffling procedures in combination with selection. For example, a first subpopulation of cells are mutagenized, allowed to recover from mutagenesis, subjected to

incomplete degradation of cell walls and then contacted with protoplasts of a second subpopulation of cells. Hybrids cells bearing markers from both subpopulations are identified (as described above) and used as the starting materials in a subsequent round of shuffling. This selection scheme selects both for cells with capacity for spontaneous protoplast formation and for cells with enhanced recombinogenicity.

In a further variation, cells having capacity for spontaneous protoplast formation can be crossed with cells having enhanced recombinogenicity evolved using other methods of the invention. The hybrid cells are particularly suitable hosts for whole genome shuffling.

Cells with mutations in enzymes involved in cell wall synthesis or maintenance can undergo fusion simply as a result of propagating the cells in osmotic-protected culture due to spontaneous protoplast formation. If the mutation is conditional, cells are shifted to a nonpermissive condition. Protoplast formation and fusion can be accelerated by addition of promoting agents, such as PEG or an electric field (See Philipova & Venkov, *Yeast* 6, 205-212 (1990); Tsoneva et al., *FEMS Microbiol. Lett.* 51, 61-65 (1989)).

5. Shuffling Methods in Yeast

Yeasts are subspecies of fungi that grow as single cells. Yeasts are used for the production of fermented beverages and leavening, for production of ethanol as a fuel, low molecular weight compounds, and for the heterologous production of proteins and enzymes (see accompanying list of yeast strains and their uses). Commonly used strains of yeast include *Saccharomyces cerevisiae, Pichia sp., Canidia sp.* and *Schizosaccharomyces pombe.*

Several types of vectors are available for cloning in yeast including integrative plasmid (YIp), yeast replicating plasmid (YRp, such as the 2μ circle based vectors), yeast episomal plasmid (YEp), yeast centromeric plasmid (YCp), or yeast artificial chromosome (YAC). Each vector can carry markers useful to select for the presence of the plasmid such as LUE2, URA3, and H1S3, or the absence of

the plasmid such as URA3 (a gene that is toxic to cells grown in the presence of 5-fluoro orotic acid.

Many yeasts have a sexual cycle and asexual (vegetative) cycles. The sexual cycle involves the recombination of the whole genome of the organism each time the cell passes through meiosis. For example, when diploid cells of *S. cerevisiae* are exposed to nitrogen and carbon limiting conditions, diploid cells undergo meiosis to form asci. Each ascus holds four haploid spores, two of mating type "a" and two of mating type "α." Upon return to rich medium, haploid spores of opposite mating type mate to form diploid cells once again. Asci of opposite mating type can mate within the ascus, or if the ascus is degraded, for example with zymolase, the haploid cells can mate with spores from other asci. This sexual cycle provides a format to shuffle endogenous genomes of yeast and/or exogenous fragment libraries inserted into yeast. This process results in swapping or accumulation of hybrid genes, and for the shuffling of homologous sequences shared by mating cells.

Yeast strains having mutations in several known genes have properties useful for shuffling. These properties include increasing the frequency of recombination and increasing the frequency of spontaneous mutations within a cell. These properties can be the result of mutation of a coding sequence or altered expression (usually overexpression) of a wildtype coding sequence. The HO nuclease effects the transposition of HMLa/α and HMRa/α to the MAT locus resulting in mating type switching. Mutants in the gene encoding this enzyme do not switch their mating type and can be employed to force crossing between strains of defined genotype, such as ones that harbor a library or have a desired phenotype and to prevent in breeding of starter strains. PMS1, MLH1, MSH2, MSH6 are involved in mismatch repair. Mutations in these genes all have a mutator phenotype (Chambers et al., *Mol. Cell. Biol.* 16, 6110-6120 (1996)). Mutations in TOP3 DNA topoisomerase have a 6-fold enhancement of interchromosomal homologous recombination (Bailis et al., *Molecular and Cellular Biology* 12, 4988-4993 (1992)). The RAD50-57 genes

confer resistance to radiation. Rad3 functions in excision of pyrimidine dimers. RAD52 functions in gene conversion. RAD50, MRE11, XRS2 function in both homologous recombination and illegitimate recombination. HOP1, RED1 function in early meiotic recombination (Mao-Draayer, *Genetics* 144, 71-86) Mutations in either HOP1 or RED1 reduce double stranded breaks at the HIS2 recombination hotspot. Strains deficient in these genes are useful for maintaining stability in hyper recombinogenic constructs such as tandem expression libraries carried on YACs. Mutations in HPR 1 are hyperrecombinogenic. HDF1 has DNA end binding activity and is involved in double stranded break repair and V(D)J recombination. Strains bearing this mutation are useful for transformation with random genomic fragments by either protoplast fusion or electroporation. Kar-1 is a dominant mutation that prevents karyogamy. Kar-1 mutants are useful for the directed transfer of single chromosomes from a donor to a recipient strain. This technique has been widely used in the transfer of YACs between strains, and is also useful in the transfer of evolved genes/chromosomes to other organisms (Markie, *YAC Protocols*, (Humana Press, Totowa, NJ, 1996). HOT1 is an *S. cerevisiae* recombination hotspot within the promoter and enhancer region of the rDNA repeat sequences. This locus induces mitotic recombination at adjacent sequences- presumably due to its high level transcription. Genes and/or pathways inserted under the transcriptional control of this region undergo increased mitotic recombination. CDC2 encodes polymerase δ and is necessary for mitotic gene conversion. Overexpression of this gene can be used in a shuffler or mutator strain. A temperature sensitive mutation in CDC4 halts the cell cycle at G1 at the restrictive temperature and could be used to synchronize protoplasts for optimized fusion and subsequent recombination.

As with filamentous fungi, the general goals of shuffling yeast include improvement in yeast as a host organism for genetic manipulation, and as a production apparatus for various compounds. One desired property in either case is to improve the capacity of yeast to express and

secrete a heterologous protein. The following example describes the use of shuffling to evolve yeast to express and secrete increased amounts of RNase A.

RNase A catalyzes the cleavage of the P-O_5, bond of RNA specifically after pyrimidine nucleotides. The enzyme is a basic 124 amino acid polypeptide that has 8 half cystine residues, each required for catalysis. YEpWL-RNase A is a vector that effects the expression and secretion of RNaseA from the yeast *S. cerevisiae*, and yeast harboring this vector secrete 1-2 mg of recombinant RNase A per liter of culture medium (delCardayré et al., *Protein Engineering* 8(3):26, 1-273 (1995)). This overall yield is poor for a protein heterologously expressed in yeast and can be improved at least 10-100 fold by shuffling. The expression of RNaseA is easily detected by several plate and microtitre plate assays (delCardayré & Raines, *Biochemistry* 33, 6031-6037 1994)). Each of the described formats for whole genome shuffling can be used to shuffle a strain of *S. cerevisiae* harboring YEpWL.RNase A, and the resulting cells can be screened for the increased secretion of RNase A into the medium. The new strains are cycled recursively through the shuffling format, until sufficiently high levels of RNase A secretion is observed. The use of RNase A is particularly useful since it not only requires proper folding and disulfide bond formation but also proper glycosylation. Thus numerous components of the expression, folding, and secretion systems can be optimized. The resulting strain is also evolved for improved secretion of other heterologous proteins.

Another goal of shuffling yeast is to increase the tolerance of yeast to ethanol. Such is useful both for the commercial production of ethanol, and for the production of more alcoholic beers and wines. The yeast strain to be shuffled acquires genetic material by exchange or transformation with other strain(s) of yeast, which may or may not be know to have superior resistance to ethanol. The strain to be evolved is shuffled and shufflants are selected for capacity to survive exposure to ethanol. Increasing concentrations of ethanol can be used in successive rounds of

shuffling. The same principles can be used to shuffle baking yeasts for improved osmotolerance.

Another desired property of shuffling yeast is capacity to grow under desired nutritional conditions. For example, it is useful to yeast to grow on cheap carbon sources such as methanol, starch, molases, cellulose, cellobiose, or xylose depending on availability. The principles of shuffling and selection are similar to those discussed for filamentous fungi.

Another desired property is capacity to produce secondary metabolites naturally produced by filamentous fungi or bacteria, Examples of such secondary metabolites are cyclosporin A, taxol, and cephalosporins. The yeast to be evolved undergoes genetic exchange or is transformed with DNA from organism(s) that produce the secondary metabolite. For example, fungi producing taxol include *Taxomyces andreanae* and *Pestalotopis microspora* (Stierle et al., *Science* 260, 214-216 (1993); Strobel et al., *Microbiol.* 142, 435-440 (1996)). DNA can also be obtained from trees that naturally produce taxol, such as *Taxus brevifolia*. DNA encoding one enzyme in the taxol pathway, taxadiene synthase, which it is believe catalyzes the commited step in taxol biosynthesis and may be rate limiting in voveral taxol production, has been cloned (Wildung & Croteau, *J. Biol. Chem.* 271, 9201-4 (1996). The DNA is then shuffled, and shufflants are screened/selected for production of the secondary metabolite. For example, taxol production can be monitored using antibodies to taxol, by mass spectroscopy or uv spectrophotometry. Alternatively, production of intermediates in taxol synthesis or enzymes in the taxol synthetic pathway can be monitored. Concetti & Ripani, *Biol. Chem. Hoppe Seyler* 375, 419-23 (1994). Other examples of secondary metabolites are polyols, amino acids, and ergosterol.

Another desired property is to increase the flocculence of yeast to facilitate separation in preparation of ethanol. Yeast can be shuffled by any of the procedures noted above with selection for shuffled yeast forming the largest clumps.

Exemplary procedure for yeast protoplasting

Protoplast preparation in yeast is reviewed by Morgan, in *Protoplasts* (Birkhauser Verlag, Basel, 1983). Fresh cells (~10^8) are washed with buffer, for example 0.1 M potassium phosphate, then resuspended in this same buffer containing a reducing agent, such as 50 mM DTT, incubated for 1 h at 30°C with gentle agitation, and then washed again with buffer to remove the reducing agent. These cells are then resuspended in buffer containing a cell wall degrading enzyme, such as Novozyme 234 (1 mg/mL), and any of a variety of osmotic stabilizers, such as sucrose, sorbitol, NaCl, KCl, $MgSO_4$, $MgCl_2$, or NH_4Cl at any of a variety of concentrations. These suspensions are then incubated at 30°C with gentle shaking (~60 rpm) until protoplasts are released. To generate protoplasts that are more likely to produce productive fusants several strategies are possible.

Protoplast formation can be increased if the cell cycle of the protoplasts have been synchronized to be halted at G1. In the case of *S. cerevisiae* this can be accomplished by the addition of mating factors, either a or α (Curran & Carter, *J. Gen. Microbiol.* 129, 1589-1591 (1983)). These peptides act as adenylate cyclase inhibitors which by decreasing the cellular level of cAMP arrest the cell cycle at G1. In addition, sex factors have been shown to induce the weakening of the cell wall in preparation for the sexual fusion of a and α cells (Crandall & Brock, *Bacteriol. Rev.* 32, 139-163 (1968); Osumi et al., *Arch. Microbiol.* 97, 27-38 (1974)). Thus in the preparation of protoplasts, cells can be treated with mating factors or other known inhibitors of adenylate cyclase, such as leflunomide or the killer toxin from *K. lactis*, to arrest them at G1 (Sugisaki et al., *Nature* 304, 464-466 (1983)). Then after fusing of the protoplasts (step 2), cAMP can be added to the regeneration medium to induce S-phase and DNA synthesis. Alternatively, yeast strains having a temperature sensitive mutation in the CDC4 gene can be used, such that cells could be synchronized and arrested at G1. After fusion cells are returned to the permissive temperature so that DNA synthesis and growth

resumes.

Once suitable protoplasts have been prepared, it is necessary to induce fusion by physical or chemical means. An equal number of protoplasts of each cell type is mixed in phosphate buffer (0.2 M, pH 5.8, 2 x 10^8 cells/mL) containing an osmotic stabilizer, for example 0.8 M NaCl, and PEG 6000 (33% w/v) and then incubated at 30°C for 5 mm while fusion occurs. Polyols, or other compounds that bind water, can be employed. The fusants are then washed and resuspended in the osmotically stabilized buffer lacking PEG, and transferred to osmotically stabilized regeneration medium on/in which the cells can be selected or screened for a desired property.

6. Shuffling Methods Using Artificial Chromosomes

Yeast artificial chromosomes (Yacs) are yeast vectors into which very large DNA fragments (e.g., 50-2000 kb) can be cloned (see, e.g., Monaco & Larin, *Trends. Biotech.* 12(7), 280-286 (1994); Ramsay, *Mol. Biotechnol.* 1(2), 181-201 1994; Huxley, *Genet. Eng.* 16, 65-91 (1994); Jakobovits, *Curr. Biol.* 4(8), 761-3 (1994); Lamb & Gearhart, *Curr. Opin. Genet. Dev.* 5(3), 342-8 (1995); Montoliu et al., *Reprod. Fertil. Dev.* 6, 577-84 (1994)). These vectors have telomeres (Tel), a centromere (Cen), an autonomously replicating sequence (ARS), and can have genes for positive (e.g., TRP1) and negative (e.g., URA3) selection. 2YACs are maintained, replicated, and segregate as other yeast chromosomes through both meiosis and mitosis thereby providing a means to expose cloned DNA to true meiotic recombination.

YACs provide a vehicle for the shuffling of libraries of large DNA fragments *in vivo*. The substrates for shuffling are typically large fragments from 20 kb to 2 Mb. The fragments can be random fragments or can be fragments known to encode a desirable property. For example, a fragment might include an operon of genes involved in production of antibiotics. Libraries can also include whole genomes or chromosomes. Viral genomes and some bacterial genomes can be cloned intact into a single YAC. In some libraries, fragments are obtained from a single organism. Other libraries include

fragment variants, as where some libraries are obtained from different individuals or species. Fragment variants can also be generated by induced mutation. Typically, genes within fragments are expressed from naturally associated regulatory sequences within yeast. However, alternatively, individual genes can be linked to yeast regulatory elements to form an expression cassette, and a concatemer of such cassettes, each containing a different gene, can be inserted into a YAC.

In some instances, fragments are incorporated into the yeast genome, and shuffling is used to evolve improved yeast strains. In other instances, fragments remain as components of YACs throughout the shuffling process, and after acquisition of a desired property within a YAC are transferred to a desired recipient cell.

a. Methods of Evolving Yeast Strains

Fragments are cloned into a YAC vector, and the resulting YAC library is transformed into competent yeast cells. Transformants containing a YAC are identified by selecting for a positive selection marker present on the YAC. The cells are allowed to recover and are then pooled. Thereafter, the cells are induced to sporulate by transferring the cells from rich medium, to nitrogen and carbon limiting medium. In the course of sporulation, cells undergo meiosis. Spores are then induced to mate by return to rich media. Optionally spores can be lysed to stimulate mating. Mating results in recombination between YACs bearing different inserts, and between YACs and natural yeast chromosomes. The latter can be promoted by irradiating spores with ultra violet light. Recombination can give rise to new phenotypes either as a result of genes expressed by fragments on the YACs or as a result of recombination with host genes, or both.

After induction of recombination between YACs and natural yeast chromosomes, YACs are often eliminated by selecting against a negative selection marker on the YACs. For example, YACs containing the marker URA3 can be selected against by propagation on media containing 5-fluro-orotic acid. Any exogenous or altered genetic material that remains

is contained within natural yeast chromosomes. Optionally, further rounds of recombination between natural yeast chromosomes can be performed after elimination of YACs. Optionally, the same or different library of YACs can be transformed into the cells, and the above steps repeated.

After elimination of YACs, yeast are then screened or selected for a desired property. The property can be a new property conferred by transferred fragments, such as production of an antibiotic. The property can also be an improved property of the yeast such as improved capacity to express or secrete an exogenous gene, improved recombinogenicity, improved stability to temperature or solvents, or other property required of commercial or research strains of yeast.

Yeast strains surviving selection/screening are then subject to a further round of recombination. Recombination can be exclusively between the chromosomes of yeast surviving selection/screening. Alternatively, a library of fragments can be introduced into the yeast cells and recombined with endogenous yeast chromosomes as before. This library of fragments can be the same or different from the library used in the previous round of transformation. YACs are eliminated as before, followed by additional rounds of recombination and/or transformation with further YAC libraries. Recombination is followed by another round of selection/screening, as above. Further rounds of recombination/screening can be performed as needed until a yeast strain has evolved to acquire the desired property.

An exemplary scheme for evolving yeast by introduction of a YAC library is shown in Fig. 10. The first part of the figure shows yeast containing an endogenous diploid genome and a YAC library of fragments representing variants of a sequence. The library is transformed into the cells to yield 100-1000 colonies per μgDNA. Most transformed yeast cells now harbor a single YAC as well as endogenous chromosomes. Meiosis is induced by growth on nitrogen and carobon limiting medium. In the course of meiosis the YACs recombine with other chromosomes in the same cell. Haploid

spores resulting from meiosis mate and regenerated diploid forms. The diploid forms now harbor recombinant chromosomes, parts of which come from endogenous chromosomes and parts from YACs. Optionally, the YACs can now be cured from the cells by selecting against a negative selection marker present on the YACS. Irrespective whether YACS are selected against, cells are then screened or selected for a desired property. Cells surviving selection/screening are transformed with another YAC library to start another shuffling cycle.

b. Method of Evolving YACs for Transfer to Recipient Strain

These methods are based in part on the fact that multiple YACs can be harbored in the same yeast cell, and YAC-YAC recombination is known to occur (Green & Olson, *Science* 250, 94-98 1990)). Inter-YAC recombination provides a format for which families of homologous genes harbored on fragments of >20 kb can be shuffled *in vivo*.

The starting population of DNA fragments show sequence similarity with each other but differ as a result of for example, induced, allelic or species diversity. Often DNA fragments are known or suspected to encode multiple genes that function in a common pathway.

The fragments are cloned into a Yac and transformed into yeast, typically with positive selection for transformants. The transformants are induced to sporulate, as a result of which chromosomes undergo meiosis. The cells are then mated. Most of the resulting diploid cells now carry two YACs each having a different insert. These are again induced to sporulate and mated. The resulting cells harbor YACs of recombined sequence. The cells can then be screened or selected for a desired property. Typically, such selection occurs in the yeast strain used for shuffling. However, if fragments being shuffled are not expressed in yeast, YACs can be isolated and transferred to an appropriate cell type in which they are expressed for screening. Examples of such properties include the synthesis or degradation of a desired compound, increased secretion of a desired gene product, or

other detectable phenotype.

Cells surviving selection/screening are subjected to successive cycles of pooling, sporulation, mating and selection/screening until the desired phenotype has been observed. Recombination can be achieved simply by transferring cells from rich medium to carbon and nitrogen limited medium to induce sporulation, and then returning the spores to rich media to induce mating. Spores can be lysed to stimulate mating.

After YACs have been evolved to encode a desired property they can be transferred to other cell types. Transfer can be isolated DNA and retransforming, protoplast fusion or electroporation. For example, transfer of YACs from yeast to mammalian cells is discussed by Monaco & Larin, *Trends in Biotechnology* 12, 280-286 (1994); Montoliu et al., *Reprod. Fertil. Dev.* 6, 577-84 (1994); Lamb et al., *Curr. Opin. Genet. Dev.* 5, 342-8 (1995).

An exemplary scheme for shuffling a YAC fragment library in yeast is shown in Fig. 11. A library of YAC fragments representing genetic variants are transformed into yeast have diploid endogenous chromosomes. The transformed yeast continue to have diploid endogenous chromosomes, plus a single YAC. The yeast are induced to undergo meiosis and sporulate. The spores contain haploid genomes, some of which contain only endogenous yeast chromosomes, and some of which contain yeast chromosomes plus a YAC. The spores are induced to mate generating diploid cells. Some of the diploid cells now contain two YAC bearing different inserts as well as diploid endogenous chromosomes. The cells are again induced to undergo meiosis and sporulate. In cells bearing two YACs, recombination occurs between the inserts, and recombinant YACs are segregated to ascocytes. Some ascoytes thus contain haploid endogenous chromosomes plus a YAC chromosome with a recombinant insert. The ascocytes mature to spores, which can mate again generating diploid cells. Some diploid cells now possess a diploid complement of endogenous chromosomes plus two recombinant YACs. These cells can then be taken through further cycles of meiosis, sporulation and mating. In each

cycle, further recombination occurs between YAC inserts and further recombinant forms of inserts are generated. After one or several cycles of recombination has occurred, cells can be tested for acquisition of a desired property. Further cycles of recombination, followed by selection, can then be performed in similar fashion.

c. Use of YACs to Clone Unlinked Genes

Shuffling of YACs is particularly amenable to transfer of unlinked but functionally related genes from one species to another, particularly where such genes have not been identified. Such is the case for several commercially important natural products, such as taxol. Transfer of the genes in the metabolic pathway to a different organism is often desirable because organisms naturally producing such compounds are not well suited for mass culturing.

Clusters of such genes can be isolated by cloning a total genomic library of DNA from an organisms producing a useful compound into a YAC library. The YAC library is then transformed into yeast. The yeast is sporulated and mated such that recombination occurs between YACs and/or between YACs and natural yeast chromosomes. Selection/screening is then performed for expression of the desired collection of genes. If the genes encode a biosynthetic pathway, expression can be detected from the appearance of product of the pathway. Production of individual enzymes in the pathway, or intermediates of the final expression product or capacity of cells to metabolize such intermediates indicates partial acquisition of the synthetic pathway. The original library or a different library can be introduced into cells surviving/selection screening, and further rounds of recombination and selection/screening can be performed until the end product of the desired metabolic pathway is produced.

7. Conjugation-Mediated Genetic Exchange

Conjugation can be employed in the evolution of cell genomes in several ways. Conjugative transfer of DNA occurs during contact between cells. See Guiney (1993) in: *Bacterial*

Conjugation (Clewell, ed., Plenum Press, New York), pp. 75-104; Reimmann & Haas *in Bacterial Conjugation* (Clewell, ed., Plenum Press, New York 1993), at pp.137-188 (incorporated by reference in their entirety for all purposes). Conjugation occurs between many types of gram negative bacteria, and some types of gram positive bacteria. Conjugative transfer is also known between bacteria and plant cells (Agrobacterium tumefaciens) or yeast. As discussed in copending application attorney docket no. 16528J-014612, the genes responsible for conjugative transfer can themselves be evolved to expand the range of cell types (e.g., from bacteria to mammals) between which such transfer can occur.

Conjugative transfer is effected by an origin of transfer (oriT) and flanking genes (MOB A, B and C), and 15-25 genes, termed tra, encoding the structures and enzymes necessary for conjugation to occur. The transfer origin is defined as the site required in cis for DNA transfer. Tra genes include tra A, B, C, D, E, F, G, H, I, J, K, L, M, N, P, Q, R, S, T, U, V, W, X, Y, Z, vir AB (alleles 1-11), C, D, E, G, IHF, and FinOP. Tra genes can be expressed in cis or trans to oriT. Other cellular enzymes, including those of the RecBCD pathway, RecA, SSB protein, DNA gyrase, DNA polI, and DNA ligase, are also involved in conjugative transfer. RecE or recF pathways can substitute for RecBCD.

One structural protein encoded by a tra gene is the sex pilus, a filament constructed of an aggregate of a single polypeptide protruding from the cell surface. The sex pilus binds to a polysaccharide on recipient cells and forms a conjugative bridge through which DNA can transfer. This process activates a site-specific nuclease encoded by a MOB gene, which specifically cleaves DNA to be transferred at oriT. The cleaved DNA is then threaded through the conjugation bridge by the action of other tra enzymes.

Mobilizable vectors can exist in episomal form or integrated into the chromosome. Episomal mobilizable vectors can be used to exchange fragments inserted into the vectors between cells. Integrated mobilizable vectors can be used to mobilize adjacent genes from the chromosome.

a. Use of Integrated Mobilizable Vectors to Promote Exchange of Genomic DNA

The F plasmid of E. coli integrates into the chromosome at high frequency and mobilizes genes unidirectional from the site of integration (Clewell, 1993, *supra;* Firth et al., *in Escherichia coli and Salmonella Cellular and Molecular Biology* 2, 2377-2401 (1996); Frost et al., *Microbiol. Rev.* 58, 162-210 (1994)). Other mobilizable vectors do not spontaneously integrate into a host chromosome at high efficiency but can be induced to do by growth under particular conditions (e.g., treatment with a mutagenic agent, growth at a nonpermissive temperature for plasmid replication). *See* Reimann & Haas *in Bacterial Conjugation* (ed. Clewell, Plenum Press, NY 1993), Ch. 6. Of particular interest is the IncP group of conjugal plasmids which are typified by their broad host range (Clewell, 1993, *supra.*

Donor "male" bacteria which bear a chromosomal insertion of a conjugal plasmid, such as the *E. coli* F factor can efficiently donate chromosomal DNA to recipient "female" enteric bacteria which lack F (F^-). Conjugal transfer from donor to recipient is initiated at *oriT*. Transfer of the nicked single strand to the recipient occurs in a 5' to 3' direction by a rolling circle mechanisms which allows mobilization of tandem chromosomal copies. Upon entering the recipient, the donor strand is discontinuously replicated. The linear, single-stranded donor DNA strand is a potent substrate for initiation of recA-mediated homologous recombination within the recipient. Recombination between the donor strand and recipient chromosomes can result in the inheritance of donor traits. Accordingly, strains which bear a chromosomal copy of F are designated Hfr (for high frequency of recombination) (Low, 1996 *in Escherichia coli and Salmonella Cellular and Molecular Biology* Vol. 2, pp. 2402-2405; Sanderson, *in Escherichia coli and Salmonella Cellular and Molecular Biology* 2, 2406-2412 (1996)).

The ability of strains with integrated mobilizable vector to transfer chromosomal DNA provides a rapid and efficient means of exchanging genetic material between a

population of bacteria thereby allowing combination of positive mutations and dilution of negative mutations. Such shuffling methods typically start with a population of strains with an integrated mobilizable vector encompassing at least some genetic diversity. The genetic diversity can be the result of natural variation, exposure to a mutagenic agent or introduction of a fragment library. The population of cells is cultured without selection to allow genetic exchange, recombination and expression of recombinant genes. The cells are then screened or selected for a evolution toward a desired property. The population surviving selection/screening can then be subject to a further round of shuffling by HFR-mediated genetic exchange or otherwise.

The natural efficiency of Hfr and other strains with integrated mob vectors as recipients of conjugal transfer can be improved by several means. The relatively low recipient efficiency of natural HFR strains is attributable to the products of *traS* and *traT* genes of F (Clewell, 1993, *supra;* Firth et al., 1996, *supra;* Frost et al., 1994, *supra;* Achtman et al., *J. Mol. Biol.* 138, 779-795 (1980). These products are localized to the inner and outer membranes of F^+ strains, respectively, where they serve to inhibit redundant matings between two strains which are both capable of donating DNA. The effects of traS and traT, and cognate genes in other strains, can be eliminated by use of knockout cells incapable of expressing these enzymes or reduced by propagating cells on a carbon-limited source. (Peters et al., *J. Bacteriol.*, 178, 3037-3043 (1996)).

In some methods, the starting population of cells has mobilizable vector integrated at different genomic sites. Directional transfer from *oriT* typically results in more frequent inheritance of traits proximal to *oriT*. This is because mating pairs are fragile and tend to dissociate (particularly when in liquid medium) resulting in the interruption of transfer. In a population of cells having mobilizable vector integrated at different sites chromosomal exchange occurs in a more random fashion. Kits of Hfr strains are available from the *E. coli*. Genetic Stock Center and the

Salmonella Genetic Stock Centre (Frost et al., 1994, *supra*). Alternatively, a library of strains with *oriT* at random sites and orientations can be produced by insertion mutagenesis using a transposon which bears *oriT*. Transfer functions for mobilization from the transposon-borne *oriT* sites could be provide by a helper vector.

Optionally, strains bearing integrated mobilizable vectors are defective in mismatch repair gene(s). Inheritance of donor traits which arise from sequence heterologies increases in strains lacking the methyl-directed mismatch repair system.

Intergenic congual transfer between species such as *E. coli* and *Salmonella typhimurium*, which are 20% divergent at the DNA level, is also possible if the recipient strain is mutH, mutL or mutS (see Rayssiguier et al., *Nature* 342, 396-401 (1989)). Such transfer can be used to obtain recombination at several points as shown by the following example.

The example uses an *S. typhimurium* Hfr donor strain having markers thr557 at map position 0, pyrF2690 at 33 min, serA13 at 62 min and hfrK5 at 43 min. MutS +/-, F- E. coli recipient strains had markers pyrD68 at 21 min aroC355 at 51 min, ilv3164 at 85 min and mutS215 at 59 min. The triauxotrophic *S. typhimurium* Hfr donor and isogenic mutS+/- triauxotrophic E. coli recipient were inoculated into 3 ml of Lb broth and shaken at 37°C until fully grown. 100 µl of the donor and each recipient were mixed in 10 ml fresh LB broth, and then deposited to a sterile Millipore 0.45 µM HA filter using a Nalgene 250 ml resuable filtration device. The donor and recipients alone were similarly diluted and deposited to check for reversion. The filters with cells were placed cell-side-up on the surface of an LB agar plate which was incubated overnight at 37°C. The filters were removed with the aid of a sterile forceps and placed in a sterile 50 ml tube containing 5 ml of minimal salts broth. Vigorous vortexing was used to wash the cells from the filters. 100 µl of mating mixtures, as well as donor and recipient controls were spread to LB for viable cell counts and minimal glucose supplemented with

either two of the three recpient requirements for single recombinant counts, one of the three requirements for double recombinant counts, or none of the three requirements for triple recombinant counts. The plates were incubated for 48 hr at 37° after which colones were counted.

Medium	Recombinant	Recombinant CFUs/Total CFUs		*mutS⁻ / mutS⁺*
Supplements	*Genotype*	*mutS⁺*	*mutS⁻*	
Aro + Iiv	pyr⁺ aro⁻ ilv⁻	-	-	-
Aro + Ura	pyr⁻ aro⁻ ilv⁺	1.2×10^{-8}	2.5×10^{-6}	208
Ilv + Ura	pyr⁻ aro⁺ ilv⁻	2.7×10^{-8}	3.0×10^{-6}	111
Aro	pyr⁺ aro⁻ ilv⁺	-	-	-
Ilv	pyr⁺ aro⁺ ilv⁻	-	-	-
Ura	pyr⁻ aro⁻ ilv⁺	$<10^{-9}$	$<10^{-9}$	
nothing	pyr⁺ aro⁺ ilv⁺			

Aro = aromatic amino acids and vitamins
Ilv = branched chain amino acids
Ura = uracil

The data indicate that recombinants can be generated at reasonable frequenceis using Hfr matings. Intergeneric recombination is enhanced 100-200 fold in a recpient that is defective methyl-directed mismatch repair.

b. Introduction of Fragments by Conjugation

Mobilizable vectors can also be used to transfer fragment libraries into cells to be evolved. This approach is particularly useful in situations in which the cells to be evolved cannot be efficiently transformed directly with the fragment library but can undergo conjugation with primary cells that can be transformed with the fragment library.

DNA fragments to be introduced into host cells encompasses diversity relative to the host cell genome. The diversity can be the result of natural diversity or mutagenesis. The DNA fragment library is cloned into a mobilizable vector having an origin of transfer. Some such vectors also contain mob genes although alternatively these functions can also be provided in trans. The vector should be

capable of efficient conjugal transfer between primary cells and the intended host cells. The vector should also confer a selectable phenotype. This phenotype can be the same as the phenotype being evolved or can be conferred by a marker, such as a drug resistance marker. The vector should preferably allow self-elimination in the intended host cells thereby allowing selection for cells in which a cloned fragment has undergone genetic exchange with a host homologous host segment rather than duplication. Such can be achieved by use of vector lacking an origin of replication functional in the intended host type or inclusion of a negative selection marker in the vector.

One suitable vector is the broad host range conjugation plasmid described by Simon et al., *Bio/Technology* 1, 784-791 (1983); TrieuCuot et al., *Gene* 102, 99-104 (1991); Bierman et al., *Gene* 116, 43-49 (1992). These plasmids can be transformed into E. coli and then force-mated into bacteria that are difficult or impossible to transform by chemical or electrical induction of competence. These plasmids contain the origin of the IncP plasmid, *oriT*. Mobilization functions are supplied in trans by chromosomally-integrated copies of the necessary genes. Conjugal transfer of DNA can in some cases be assisted by treatment of the recipient (if gram-positive) with sub-inhibitory concentrations of penicillins (Trieu-Cuot et al., 1993 *FEMS Microbiol. Lett.* 109, 19-23).

Cells that have undergone allelic exchange with library fragments can be screened or selected for evolution toward a desired phenotype. Subsequent rounds of recombination can be performed by repeating the conjugal transfer step. the library of fragments can be fresh or can be obtained from some (but not all) of the cells surviving a previous round of selection/screening. Conjugation-mediated shuffling can be combined with other methods of shuffling.

8. Genetic Exchange Promoted by Transducing Phage

Transduction is the transfer, from one cell to another, of nonviral genetic material within a viral coat (Masters, *in Escherichia coli and Salmonella Cellular and*

Molecular Biology 2, 2421-2442 (1996). Perhaps the two best examples of generalized transducing phage are bacteriophages P1 and P22 of *E. coli* and *S. typhimurium*, respectively. Generalized transducing bacteriophage particles are formed at a low frequency during lytic infection when viral-genome-sized, doubled-stranded fragments of host (which serves as donor) chromosomal DNA are packaged into phage heads. Promiscuous high transducing (HT) mutants of bacteriophage P22 which efficiently package DNA with little sequence specificity have been isolated. Infection of a susceptible host results in a lysate in which up to 50% of the phage are transducing particles. Adsorption of the generalized transducing particle to a susceptible recipient cell results in the injection of the donor chromosomal fragment. RecA-mediated homologous recombination following injection of the donor fragment can result in the inheritance of donor traits.

Generalized transducing phage can be used to exchange genetic material between a population of cells encompassing genetic diversity and susceptible to infection by the phage. Genetic diversity can be the result of natural variation between cells, induced mutation of cells or the introduction of fragment libraries into cells. DNA is then exchanged between cells by generalized transduction. If the phage does not cause lysis of cells, the entire population of cells can be propagated in the presence of phage. If the phage results in lytic infection, transduction is performed on a split pool basis. That is, the starting population of cells is divided into two. One subpopulation is used to prepare transducing phage. The transducing phage are then infected into the other subpopulation. Preferably, infection is performed at high multiplicity of phage per cell so that few cells remain uninfected. Cells surviving infection are propagated and screened or selected for evolution toward a desired property. The pool of cells surviving screening/selection can then be shuffled by a further round of generalized transduction or by other shuffling methods.

The efficiency of the above methods can be increased by reducing infection of cells by infectious (nontransducing

phage) and by reducing lysogen formation. The former can be achieved by inclusion of chelators of divalent cations, such as citrate and EGTA in culture media. Divalent cations are required for phage absorption and the inclusion of chelating agents therefore provides a means of preventing unwanted infection. Integration defective (*int*⁻) derivatives of generalized transducing phage can be used to prevent lysogen formation. In a further variation, host cells with defects in mismatch repair gene(s) can be used to increase recombination between transduced DNA and genomic DNA.

V. Methods for Recursive Sequence Recombination

Some formats and examples for recursive sequence recombination, sometimes referred to as DNA shuffling or molecular breeding, have been described by the present inventors and co-workers in copending application, attorney docket no. 16528A-014612, filed March 25, 1996, PCT/US95/02126 filed February 17, 1995 (published as WO 95/22625); Stemmer, Science 270, 1510 (1995); Stemmer et al., *Gene*, 164, 49-53 (1995); Stemmer, *Bio/Technology*, 13, 549-553 (1995); Stemmer, *Proc. Natl. Acad. Sci. USA* 91, 10747-10751 (1994); Stemmer, *Nature* 370, 389-391 (1994); Crameri et al., *Nature Medicine*, 2(1):1-3, (1996); Crameri et al., *Nature Biotechnology* 14, 315-319 (1996) (each of which is incorporated by reference in its entirety for all purposes).

(1) *In Vitro* Formats

One format for shuffling *in vitro* is illustrated in Fig. 1. The initial substrates for recombination are a pool of related sequences. The X's in the Fig. 1, panel A, show where the sequences diverge. The sequences can be DNA or RNA and can be of various lengths depending on the size of the gene or DNA fragment to be recombined or reassembled. Preferably the sequences are from 50 bp to 50 kb.

The pool of related substrates are converted into overlapping fragments, e.g., from about 5 bp to 5 kb or more, as shown in Fig. 1, panel B. Often, the size of the fragments is from about 10 bp to 1000 bp, and sometimes the size of the DNA fragments is from about 100 bp to 500 bp. The conversion

can be effected by a number of different methods, such as DNAseI or RNAse digestion, random shearing or partial restriction enzyme digestion. Alternatively, the conversion of substrates to fragments can be effected by incomplete PCR amplification of substrates or PCR primed from a single primer. Alternatively, appropriate single-stranded fragments can be generated on a nucleic acid synthesizer. The concentration of nucleic acid fragments of a particular length and sequence is often less than 0.1 % or 1% by weight of the total nucleic acid. The number of different specific nucleic acid fragments in the mixture is usually at least about 100, 500 or 1000.

The mixed population of nucleic acid fragments are converted to at least partially single-stranded form. Conversion can be effected by heating to about 80 °C to 100 °C, more preferably from 90 °C to 96 °C, to form single-stranded nucleic acid fragments and then reannealing. Conversion can also be effected by treatment with single-stranded DNA binding protein or recA protein. Single-stranded nucleic acid fragments having regions of sequence identity with other single-stranded nucleic acid fragments can then be reannealed by cooling to 20 °C to 75 °C, and preferably from 40 °C to 65 °C. Renaturation can be accelerated by the addition of polyethylene glycol (PEG), other volume-excluding reagents or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt concentration is from 10 mM to 100 mM. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%. The fragments that reanneal can be from different substrates as shown in Fig. 1, panel C. The annealed nucleic acid fragments are incubated in the presence of a nucleic acid polymerase, such as Taq or Klenow, or proofreading polymerases, such as pfu or pwo, and dNTP's (i.e. dATP, dCTP, dGTP and dTTP). If regions of sequence identity are large, Taq polymerase can be used with an annealing temperature of between 45-65°C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30°C (Stemmer, *Proc. Natl. Acad.*

Sci. USA (1994), *supra*). The polymerase can be added to the random nucleic acid fragments prior to annealing, simultaneously with annealing or after annealing.

The process of denaturation, renaturation and incubation in the presence of polymerase of overlapping fragments to generate a collection of polynucleotides containing different permutations of fragments is sometimes referred to as shuffling of the nucleic acid *in vitro*. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from 2 to 100 times, more preferably the sequence is repeated from 10 to 40 times. The resulting nucleic acids are a family of double-stranded polynucleotides of from about 50 bp to about 100 kb, preferably from 500 bp to 50 kb, as shown in Fig. 1, panel D. The population represents variants of the starting substrates showing substantial sequence identity thereto but also diverging at several positions. The population has many more members than the starting substrates. The population of fragments resulting from shuffling is used to transform host cells, optionally after cloning into a vector.

In a variation of *in vitro* shuffling, subsequences of recombination substrates can be generated by amplifying the full-length sequences under conditions which produce a substantial fraction, typically at least 20 percent or more, of incompletely extended amplification products. The amplification products, including the incompletely extended amplification products are denatured and subjected to at least one additional cycle of reannealing and amplification. This variation, in which at least one cycle of reannealing and amplification provides a substantial fraction of incompletely extended products, is termed "stuttering." In the subsequent amplification round, the incompletely extended products reanneal to and prime extension on different sequence-related template species.

In a further variation, a mixture of fragments is spiked with one or more oligonucleotides. The oligonucleotides can be designed to include precharacterized mutations of a wildtype sequence, or sites of natural

variations between individuals or species. The oligonucleotides also include sufficient sequence or structural homology flanking such mutations or variations to allow annealing with the wildtype fragments. Some oligonucleotides may be random sequences. Annealing temperatures can be adjusted depending on the length of homology.

In a further variation, recombination occurs in at least one cycle by template switching, such as when a DNA fragment derived from one template primes on the homologous position of a related but different template. Template switching can be induced by addition of recA, rad51, rad55, rad57 or other polymerases (e.g., viral polymerases, reverse transcriptase) to the amplification mixture. Template switching can also be increased by increasing the DNA template concentration.

In a further variation, at least one cycle of amplification can be conducted using a collection of overlapping single-stranded DNA fragments of related sequence, and different lengths. Fragments can be prepared using a single stranded DNA phage, such as M13. Each fragment can hybridize to and prime polynucleotide chain extension of a second fragment from the collection, thus forming sequence-recombined polynucleotides. In a further variation, ssDNA fragments of variable length can be generated from a single primer by Vent or other DNA polymerase on a first DNA template. The single stranded DNA fragments are used as primers for a second, Kunkel-type template, consisting of a uracil-containing circular ssDNA. This results in multiple substitutions of the first template into the second. *See* Levichkin et al., *Mol. Biology* 29, 572-577 (1995).

(2) *In Vivo* Formats

(a) Plasmid-Plasmid Recombination

The initial substrates for recombination are a collection of polynucleotides comprising variant forms of a gene. The variant forms often show substantial sequence identity to each other sufficient to allow homologous

recombination between substrates. The diversity between the polynucleotides can be natural (e.g., allelic or species variants), induced (e.g., error-prone PCR), or the result of *in vitro* recombination. Diversity can also result from resynthesizing genes encoding natural proteins with alternative and/or mixed codon usage. There should be at least sufficient diversity between substrates that recombination can generate more diverse products than there are starting materials. There must be at least two substrates differing in at least two positions. However, commonly a library of substrates of 10^3-10^8 members is employed. The degree of diversity depends on the length of the substrate being recombined and the extent of the functional change to be evolved. Diversity at between 0.1-50% of positions is typical. The diverse substrates are incorporated into plasmids. The plasmids are often standard cloning vectors, e.g., bacterial multicopy plasmids. However, in some methods to be described below, the plasmids include mobilization functions. The substrates can be incorporated into the same or different plasmids. Often at least two different types of plasmid having different types of selection marker are used to allow selection for cells containing at least two types of vector. Also, where different types of plasmid are employed, the different plasmids can come from two distinct incompatibility groups to allow stable co-existence of two different plasmids within the cell. Nevertheless, plasmids from the same incompatibility group can still co-exist within the same cell for sufficient time to allow homologous recombination to occur.

Plasmids containing diverse substrates are initially introduced into procaryotic or eucaryotic cells by any transfection methods (e.g., chemical transformation, natural competence, electroporation, viral transduction or biolistics). Often, the plasmids are present at or near saturating concentration (with respect to maximum transfection capacity) to increase the probability of more than one plasmid entering the same cell. The plasmids containing the various substrates can be transfected simultaneously or in multiple

rounds. For example, in the latter approach cells can be transfected with a first aliquot of plasmid, transfectants selected and propagated, and then infected with a second aliquot of plasmid.

Having introduced the plasmids into cells, recombination between substrates to generate recombinant genes occurs within cells containing multiple different plasmids merely by propagating in the cells. However, cells that receive only one plasmid are unable to participate in recombination and the potential contribution of substrates on such plasmids to evolution is not fully exploited (although these plasmids may contribute to some extent if they are propagated in mutator cells or otherwise accumulate point mutations (i.e., by ultraviolet radiation treatment). The rate of evolution can be increased by allowing all substrates to participate in recombination. Such can be achieved by subjecting transfected cells to electroporation. The conditions for electroporation are the same as those conventionally used for introducing exogenous DNA into cells (e.g., 1,000-2,500 volts, 400 μF and a 1-2 mM gap). Under these conditions, plasmids are exchanged between cells allowing all substrates to participate in recombination. In addition the products of recombination can undergo further rounds of recombination with each other or with the original substrate. The rate of evolution can also be increased by use of conjugative transfer. Conjugative transfer systems are known in many bacteria (*E. coli*, *P. aeruginosa*, *S. pneumoniae*, and *H. influenzae*) and can also be used to transfer DNA between bacteria and yeast or between bacteria and mammalian cells.

To exploit conjugative transfer, substrates are cloned into plasmids having MOB genes, and *tra* genes are also provided in cis or in trans to the MOB genes. The effect of conjugative transfer is very similar to electroporation in that it allows plasmids to move between cells and allows recombination between any substrate and the products of previous recombination to occur merely by propagating the culture. The details of how conjugative transfer is exploited

in these vectors are discussed in more detail below. The rate of evolution can also be increased by fusing protoplasts of cells to induce exchange of plasmids or chromosomes. Fusion can be induced by chemical agents, such as PEG, or viruses or viral proteins, such as influenza virus hemagglutinin, HSV-1 gB and gD. The rate of evolution can also be increased by use of mutator host cells (e.g., Mut L, S, D, T, H and *Ataxia telangiectasia* human cell lines).

The time for which cells are propagated and recombination is allowed to occur, of course, varies with the cell type but is generally not critical, because even a small degree of recombination can substantially increase diversity relative to the starting materials. Cells bearing plasmids containing recombined genes are subject to screening or selection for a desired function. For example, if the substrate being evolved contains a drug resistance gene, one selects for drug resistance. Cells surviving screening or selection can be subjected to one or more rounds of screening/selection followed by recombination or can be subjected directly to an additional round of recombination.

The next round of recombination can be achieved by several different formats independently of the previous round. For example, a further round of recombination can be effected simply by resuming the electroporation or conjugation-mediated intercellular transfer of plasmids described above. Alternatively, a fresh substrate or substrates, the same or different from previous substrates, can be transfected into cells surviving selection/screening. Optionally, the new substrates are included in plasmid vectors bearing a different selective marker and/or from a different incompatibility group than the original plasmids. As a further alternative, cells surviving selection/screening can be subdivided into two subpopulations, and plasmid DNA from one subpopulation transfected into the other, where the substrates from the plasmids from the two subpopulations undergo a further round of recombination. In either of the latter two options, the rate of evolution can be increased by employing DNA extraction, electroporation, conjugation or mutator cells, as

described above. In a still further variation, DNA from cells surviving screening/selection can be extracted and subjected to *in vitro* DNA shuffling.

After the second round of recombination, a second round of screening/selection is performed, preferably under conditions of increased stringency. If desired, further rounds of recombination and selection/screening can be performed using the same strategy as for the second round. With successive rounds of recombination and selection/screening, the surviving recombined substrates evolve toward acquisition of a desired phenotype. Typically, in this and other methods of recursive recombination, the final product of recombination that has acquired the desired phenotype differs from starting substrates at 0.1%-25% of positions and has evolved at a rate orders of magnitude in excess (e.g., by at least 10-fold, 100-fold, 1000-fold, or 10,000 fold) of the rate of naturally acquired mutation of about 1 mutation per 10^{-9} positions per generation (*see* Anderson & Hughes, *Proc. Natl. Acad. Sci. USA* 93, 906-907 (1996)).

(b) Virus-Plasmid Recombination

The strategy used for plasmid-plasmid recombination can also be used for virus-plasmid recombination; usually, phage-plasmid recombination. However, some additional comments particular to the use of viruses are appropriate. The initial substrates for recombination are cloned into both plasmid and viral vectors. It is usually not critical which substrate(s) are inserted into the viral vector and which into the plasmid, although usually the viral vector should contain different substrate(s) from the plasmid. As before, the plasmid (and the virus) typically contains a selective marker. The plasmid and viral vectors can both be introduced into cells by transfection as described above. However, a more efficient procedure is to transfect the cells with plasmid, select transfectants and infect the transfectants with virus. Because the efficiency of infection of many viruses approaches 100% of cells, most cells transfected and infected by this

route contain both a plasmid and virus bearing different substrates.

Homologous recombination occurs between plasmid and virus generating both recombined plasmids and recombined virus. For some viruses, such as filamentous phage, in which intracellular DNA exists in both double-stranded and single-stranded forms, both can participate in recombination. Provided that the virus is not one that rapidly kills cells, recombination can be augmented by use of electroporation or conjugation to transfer plasmids between cells. Recombination can also be augmented for some types of virus by allowing the progeny virus from one cell to reinfect other cells. For some types of virus, virus infected-cells show resistance to superinfection. However, such resistance can be overcome by infecting at high multiplicity and/or using mutant strains of the virus in which resistance to superinfection is reduced.

The result of infecting plasmid-containing cells with virus depends on the nature of the virus. Some viruses, such as filamentous phage, stably exist with a plasmid in the cell and also extrude progeny phage from the cell. Other viruses, such as lambda having a cosmid genome, stably exist in a cell like plasmids without producing progeny virions. Other viruses, such as the T-phage and lytic lambda, undergo recombination with the plasmid but ultimately kill the host cell and destroy plasmid DNA. For viruses that infect cells without killing the host, cells containing recombinant plasmids and virus can be screened/selected using the same approach as for plasmid-plasmid recombination. Progeny virus extruded by cells surviving selection/screening can also be collected and used as substrates in subsequent rounds of recombination. For viruses that kill their host cells, recombinant genes resulting from recombination reside only in the progeny virus. If the screening or selective assay requires expression of recombinant genes in a cell, the recombinant genes should be transferred from the progeny virus to another vector, e.g., a plasmid vector, and retransfected into cells before selection/screening is performed.

For filamentous phage, the products of recombination

are present in both cells surviving recombination and in phage extruded from these cells. The dual source of recombinant products provides some additional options relative to the plasmid-plasmid recombination. For example, DNA can be isolated from phage particles for use in a round of *in vitro* recombination. Alternatively, the progeny phage can be used to transfect or infect cells surviving a previous round of screening/selection, or fresh cells transfected with fresh substrates for recombination.

(c) Virus-Virus Recombination

The principles described for plasmid-plasmid and plasmid-viral recombination can be applied to virus-virus recombination with a few modifications. The initial substrates for recombination are cloned into a viral vector. Usually, the same vector is used for all substrates. Preferably, the virus is one that, naturally or as a result of mutation, does not kill cells. After insertion, some viral genomes can be packaged *in vitro*. The packaged viruses are used to infect cells at high multiplicity such that there is a high probability that a cell receives multiple viruses bearing different substrates.

After the initial round of infection, subsequent steps depend on the nature of infection as discussed in the previous section. For example, if the viruses have phagemid genomes such as lambda cosmids or M13, F1 or Fd phagemids, the phagemids behave as plasmids within the cell and undergo recombination simply by propagating the cells. Recombination can be augmented by electroporation of cells. Following selection/screening, cosmids containing recombinant genes can be recovered from surviving cells (e.g., by heat induction of a cos^- lysogenic host cell), repackaged *in vitro*, and used to infect fresh cells at high multiplicity for a further round of recombination.

If the viruses are filamentous phage, recombination of replicating form DNA occurs by propagating the culture of infected cells. Selection/screening identifies colonies of cells containing viral vectors having recombinant genes with

improved properties, together with phage extruded from such cells. Subsequent options are essentially the same as for plasmid-viral recombination.

(d) Chromosome-Plasmid Recombination

This format can be used to evolve both the chromosomal and plasmid-borne substrates. The format is particularly useful in situations in which many chromosomal genes contribute to a phenotype or one does not know the exact location of the chromosomal gene(s) to be evolved. The initial substrates for recombination are cloned into a plasmid vector. If the chromosomal gene(s) to be evolved are known, the substrates constitute a family of sequences showing a high degree of sequence identity but some divergence from the chromosomal gene. If the chromosomal genes to be evolved have not been located, the initial substrates usually constitute a library of DNA segments of which only a small number show sequence identity to the gene or gene(s) to be evolved. Divergence between plasmid-borne substrate and the chromosomal gene(s) can be induced by mutagenesis or by obtaining the plasmid-borne substrates from a different species than that of the cells bearing the chromosome.

The plasmids bearing substrates for recombination are transfected into cells having chromosomal gene(s) to be evolved. Evolution can occur simply by propagating the culture, and can be accelerated by transferring plasmids between cells by conjugation or electroporation. Evolution can be further accelerated by use of mutator host cells or by seeding a culture of nonmutator host cells being evolved with mutator host cells and inducing intercellular transfer of plasmids by electroporation or conjugation. Preferably, mutator host cells used for seeding contain a negative selection marker to facilitate isolation of a pure culture of the nonmutator cells being evolved. Selection/screening identifies cells bearing chromosomes and/or plasmids that have evolved toward acquisition of a desired function.

Subsequent rounds of recombination and selection/screening proceed in similar fashion to those

described for plasmid-plasmid recombination. For example, further recombination can be effected by propagating cells surviving recombination in combination with electroporation or conjugative transfer of plasmids. Alternatively, plasmids bearing additional substrates for recombination can be introduced into the surviving cells. Preferably, such plasmids are from a different incompatibility group and bear a different selective marker than the original plasmids to allow selection for cells containing at least two different plasmids. As a further alternative, plasmid and/or chromosomal DNA can be isolated from a subpopulation of surviving cells and transfected into a second subpopulation. Chromosomal DNA can be cloned into a plasmid vector before transfection.

(e) Virus-Chromosome Recombination

As in the other methods described above, the virus is usually one that does not kill the cells, and is often a phage or phagemid. The procedure is substantially the same as for plasmid-chromosome recombination. Substrates for recombination are cloned into the vector. Vectors including the substrates can then be transfected into cells or *in vitro* packaged and introduced into cells by infection. Viral genomes recombine with host chromosomes merely by propagating a culture. Evolution can be accelerated by allowing intercellular transfer of viral genomes by electroporation, or reinfection of cells by progeny virions. Screening/selection identifies cells having chromosomes and/or viral genomes that have evolved toward acquisition of a desired function.

There are several options for subsequent rounds of recombination. For example, viral genomes can be transferred between cells surviving selection/recombination by electroporation. Alternatively, viruses extruded from cells surviving selection/screening can be pooled and used to superinfect the cells at high multiplicity. Alternatively, fresh substrates for recombination can be introduced into the cells, either on plasmid or viral vectors.

EXAMPLES

1. Evolving Hyper-Recombinogenic RecA

RecA protein is implicated in most *E. coli* homologous recombination pathways. Most mutations in recA inhibit recombination, but some have been reported to increase recombination (Kowalczykowski et al., *Microbiol. Rev.*, 58, 401-465 (1994)). The following example describes evolution of RecA to acquire hyper-recombinogenic activity useful in *in* vivo shuffling formats.

Hyperrecombinogenic RecA was selected using a modification of a system developed by Shen et al., *Genetics* 112, 441-457 (1986); Shen et al., *Mol. Gen. Genet.* 218, 358-360 (1989)) to measure the effect of substrate length and homology on recombination frequency. Shen & Huang's system used plasmids and bacteriophages with small (31-430 bp) regions of homology at which the two could recombine. In a restrictive host, only phage that had incorporated the plasmid sequence were able to form plaques.

For shuffling of recA, endogenous recA and mutS were deleted from host strain MC1061. In this strain, no recombination was seen between plasmid and phage. *E. coli* recA was then cloned into two of the recombination vectors (Bp221 and πMT631c18). Plasmids containing cloned RecA were able to recombine with homologous phage:λV3 (430 bp identity with Bp221),λV13 (430 bp stretch of 89% identity with Bp221) and λlink H (31bp identity with πMt631c18, except for 1 mismatch at position 18).

The cloned RecA was then shuffled *in vitro* using the standard DNase-treatment followed by PCR-based reassembly. Shuffled plasmids were transformed into the nonrecombining host strain. These cells were grown up overnight, infected with phage λVc, λV13 or λlink H, and plated onto NZCYM plates in the presence of a 10-fold excess of MC1061 lacking plasmid. The more efficiently a recA allele promotes recombination between plasmid and phage, the more highly the allele is represented in the bacteriophage DNA. Consequently harvesting all the phage from the plates and recovering the recA genes selects for the most recombinogenic recA alleles.

Recombination frequencies for wild type and a pool of hyper-recombinogenic RecA after 3 rounds of shuffling were as follows:

Cross	Wild Type	Hyper Recom
BP221 x V3	6.5×10^{-4}	3.3×10^{-2}
BP221 x V13	2.2×10^{-5}	1.0×10^{-3}
$^{\pi}$MT631c18 x link H	8.7×10^{-6}	4.7×10^{-6}

These results indicate a 50-fold increase in recombination for the 430 bp substrate, and a 5-fold increase for the 31 bp substrate.

The recombination frequency between BP221 and V3 for five individual clonal isolates are shown below, and the DNA and protein sequences and alignments thereof are included in Figs. 12 and 13.

Wildtype: 1.6×10^{-4}
Clone 2: 9.8×10^{-3} (61 x increase)
Clone 4: 9.9×10^{-3} (62 x increase)
Clone 5: 6.2×10^{-3} (39 x increase)
Clone 6: 8.5×10^{-3} (53 x increase)
Clone 13: 0.019 (116 x increase)

Clones 2, 4, 5, 6 and 13 can be used as the substrates in subsequent rounds of shuffling, if further improvement in recA is desired. Not all of the variations from the wildtype recA sequence necessarily contribute to the hyperrecombinogenic phenotype. Silent variations can be eliminated by backcrossing. Alternatively, variants of recA incorporating individual points of variation from wildtype at codons 5, 18, 156, 190, 236, 268, 271, 283, 304, 312, 317, 345 and 353 can be tested for activity.

2. Whole Organism Evolution for Hyper-Recombination

The possibility of selection for an *E. coli* strain with an increased level of recombination was indicated from phenotypes of wild-type, Δ*recA*, *mutS* and Δ*recA mutS* strains following exposure to mitomycin C-an inter-strand cross-linking agent of DNA.

Exposure of *E. coli* to mitomycin C causes inter-strand cross-linking of DNA thereby blocking DNA replication. Repair of the inter-strand DNA cross links in *E. coli* occurs

via a RecA-dependent recombinational repair pathway (Friedberg et al., *in DNA Repair and Mutagenesis* (1995) pp. 191-232). Processing of cross-links during repair results in occasional double-strand DNA breaks, which too are repaired by a RecA-dependent recombinational route. Accordingly, recA$^-$ strains are significantly more sensitive than wildtype strains to mitomycin C exposure. In fact, mitomycin C is used in simple disk-sensitivity assays to differentiate between RecA$^+$ and RecA$^-$ strains.

In addition to its recombinogenic properties, mitomycin C is a mutagen. Exposure to DNA damaging agents, such as mitomycin C, typically results in the induction of the *E. coli* SOS regulon which includes products involved in error-prone repair of DNA damage (Friedberg et al., 1995, *supra*, at pp. 465-522).

Following phage P1-mediated generalized transduction of the Δ(*recA-srl*)::Tn10 allele (a nonfunctional allele) into wild-type and *mutS E. coli*, tetracycline-resistant transductants were screened for a *recA*$^-$phenotype using the mitomycin C-sensitivity assay. It was observed in LB overlays with a 1/4 inch filter disk saturated with 10 μg of mitomycin C following 48 hours at 37°C, growth of the wild-type and *mutS* strains was inhibited within a region with a radius of about 10 mm from the center of the disk. DNA cross-linking at high levels of mitomycin C saturates recombinational repair resulting in lethal blockage of DNA replication. Both strains gave rise to occasional colony forming units within the zone of inhibition, although, the frequency of colonies was ~10-20-fold higher in the *mutS* strain. This is presumably due to the increased rate of spontaneous mutation of *mutS* backgrounds. A side-by-side comparison demonstrated that the Δ*recA* and Δ*recA* *mutS* strains were significantly more sensitive to mitomycin C with growth inhibited in a region extending about 15 mm from the center of the disk. However, in contrast to the recA$^+$ strains, no Mit^r individuals were seen within the region of growth inhibition-not even in the *mutS* background. The appearance of Mit^r individuals in *recA*$^+$ backgrounds, but not in Δ*recA* backgrounds indicates the Mit^r is dependent upon a

functional RecA protein and suggests that Mit^r may result from an increased capacity for recombinational repair of mitomycin C-induced damage.

Mutations which lead to increased capacity for RecA-mediated recombinational repair may be diverse, unexpected, unlinked, and potentially synergistic. A recursive protocol alternating selection for Mit^r and chromosomal shuffling evolves individual cells with a dramatically increased capacity for recombination.

The recursive protocol is as follows. Following exposure of a *mutS* strain to mitomycin C, Mit^r individuals are pooled and cross-bread (e.g., via Hfr-mediated chromosomal shuffling or split-pool generalized transduction). Alleles which result in Mit^r and presumably result in an increased capacity for recombinational repair are shuffled among the population in the absence of mismatch repair. In addition, error-prone repair following exposure to mitomycin C can introduce new mutations for the next round of shuffling. The process is repeated using increasingly more stringent exposures to mitomycin C. A number of parallel selections in the first round as a means of generating a variety of alleles. Optionally, recombinogencity of isolates can be monitored for hyper-recombination using a plasmid x plasmid assay or a chromosome x chromosome assay (e.g., that of Konrad, *J. Bacteriol.* 130, 167-172 (1977)).

3. Whole Genome Shuffling of *Streptomyces coelicolor*

To demonstrate that recursive mutation and recombination of an entire genome can be used to improve a particular phenotype, *S. coelicolor* is being recursively shuffled both alone and with its close relative *S. lividans* to improve the overall production of the blue pigment γ-actinorhodin. This strain improvement strategy is being compared to a similar strain improvement program that does not include recombination.

Spore suspensions of *S. coelicolor* and *S. lividans* are resuspended in sterile water and subjected to UV mutagenesis (600 "energy" units) using a Stratalinker

(Stratagene), and the resulting mutants are "grown out" on sporulation agar. Spores are collected and plated on solid RG-2 medium (Bystrykh et al., *J. Bact.* 178, 2238-2244 (1996)). Colonies producing larger or darker halos of blue pigment are selected, grown in liquid RG-2 medium, and the amount of γ-actinorhodin produced is determined spectrophotometrically by measuring the absorbance at 650 nm of alkaline culture supernatents using a microtitreplate format. Pigment concentration and structure is further confirmed by LC/MS, MS/MS and/or NMR. Cells producing γ-actinorhodin at levels higher than that of wildtype are carried forward in the strain improvement program. Spores isolated from each of the mutants are either 1) again mutagenized and screened as above (no recombination control) or, 2) grown up, prepared as protoplasts, and fused. Procedures for preparing and fusing *Streptomyces* protoplasts are described in *Genetic Manipulation of Streptomyces--A Laboratory Manual*, (Hopwood, D.A. et al.). The regenerated fused protoplasts are then screened as above for clones having undergone recombination that produce γ-actinorhodin at levels higher than the cells that were fused. The selected clones are subjected to UV mutagenesis again and the screening and recombination are repeated recursively until the desired level of γ-actinorhodin production is reached.

The foregoing description of the preferred embodiments of the present invention has been presented for purposes of illustration and description. They are not intended to be exhaustive or to limit the invention to the precise form disclosed, and many modifications and variations are possible in light of the above teaching. Such modifications and variations which may be apparent to a person skilled in the art are intended to be within the scope of this invention. All patent documents and publications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each item were so individually denoted.

WHAT IS CLAIMED:

1. A method of evolving a cell to acquire a desired function, comprising:

(1) introducing a library of DNA fragments into a plurality of cells whereby at least one of the fragments undergoes recombination with a segment in the genome or an episome of the cells to produce modified cells;

(2) screening the modified cells for modified cells that have evolved toward acquisition of the desired function;

(3) recombining DNA from the modified cells that have evolved toward the desired function with a further library of DNA fragments at least one of which undergoes recombination with a segment in the genome or the episome of the modified cells to produce further modified cells;

(4) screening the further modified cells for further modified cells that have further evolved toward acquisition of the desired function;

(5) repeating (3) and (4) as required until the further modified cells have acquired the desired function.

General Methods

2. The method of claim 1, wherein the library of DNA fragments is a substantially complete genomic library from at least one heterologous cell type.

3. The method of claim 1, wherein the library of fragments comprises natural variants of a gene from different individuals.

4. The method of claim 1, further comprising subdividing the modified cells into first and second pools, isolating the further library of DNA fragments from the second pool and introducing the further library of DNA fragments into the first pool.

5. The method of claim 1, wherein the library of DNA fragments are components of viruses and the introducing

occurs by infection of the cells with the viruses.

6. The method of claim 1, wherein the library of DNA fragments is cloned into a suicide vector incapable of permanent episomal existence in the cells.

7. The method of claim 6, wherein the suicide vector further comprises a selective marker.

recA

8. The method of claim 1, further comprising coating the library or further library of DNA fragments with recA protein to stimulate recombination with the segment of the genome.

Mutant Selection

9. The method of claim 1, further comprising denaturing the library of fragments to produce single-stranded DNA, reannealling the single-stranded DNA to produce duplexes some of which contain mismatches at points of variation in the fragments, and selecting duplexes containing mismatches by affinity chromatography to immobilized *MutS*.

10. The method of claim 10, further comprising fragmenting the library of fragments to produce subfragments before denaturation, and reassembling duplexes of subfragments containing mismatches into reassembled fragments.

11. The method of claim 10, wherein the average diversity between reassembled fragments is at least five times greater than the average diversity between fragments.

Secretion

12. The method of claim 1, wherein the desired function is secretion of a protein, and the plurality of cells further comprises a construct encoding the protein.

13. The method of claim 12, wherein the protein is toxic to the plurality of cells unless secreted, and the modified or further modified cells having evolved toward acquisition of the desired function are screened by propagating the cells and recovering surviving cells.

14. The method of claim 13, wherein the protein is β-lactamase or alkaline phosphatase, and the modified or further modified cells having evolved toward acquisition of the desired function are screened by monitoring metabolism of a chromogenic substrate of the b-lactamase or alkaline phosphatase.

15. The method of claim 14, wherein the protein is an antibody and the plurality of cells is *E. coli*.

16. The method of claim 15, wherein the construct further encodes a marker which is expressed with the protein as a fusion protein, and the screening comprises propagating the modified or further modified cells and identifying cells secreting the fusion protein by FACS™ sorting.

17. The method of claim 16, wherein the marker protein is linked to a phospholipid anchoring domain that anchors the marker protein to the cell surface after secretion from the cell.

18. The method of claim 16, wherein the cells are contained in agar drops which confine secreted protein in proximity with the cell secreting the protein.

19. The method of claim 12, wherein at least one fragment in the library encodes a signal sequence, and the at least one fragment is incorporated into a construct operably linked to a sequence encoding a protein to be secreted from the cells.

20. The method of claim 12, wherein at least one

fragment in the library encodes a signal processing enzyme and the cells contain a construct encoding a protein to be secreted operably linked to a signal sequence.

21. The method of claim 12, wherein at least one fragment in the library encodes a gene selected from the group consisting of *Sec*A, *Sec*B, *Sec*E, *Sec*D and *Sec*F genes.

Recombination

22. The method of claim 1, wherein the desired function is enhanced recombination.

23. The method of claim 1, wherein the library of fragments comprises a cluster of genes collectively conferring recombination capacity.

24. The method of claim 1, wherein the at least one gene is selected from the group consisting of *rec*A, *rec*BCD, *rec*BC, *rec*E, *rec*F, *rec*G, *rec*O, *rec*Q, *rec*R, *rec*T, *ruv*A, *ruv*B, *ruv*C, *sbc*B, *ssb*, *top*A, *gyr*A and B, *lig*, *pol*A, *uvr*D, E, *rec*L, *mut*U, and *hel*D.

25. The method of claim 24, wherein the plurality of cells further comprises a gene encoding a marker whose expression is prevented by a mutation removable by recombination, and the modified or further modified cells are screened by their expression of the marker resulting from removal of the mutation by recombination.

26. The method of claim 24, wherein in the screening steps, the modified or further modified cells are exposed to a mutagen and modified or further modified cells having evolved toward acquisition of the desired function are selected by their survival of the exposure, survival being conferred by the cells' enhanced recombinational capacity to remove damage induced by the mutagen.

27. The method of claim 26, wherein the mutagen is

radiation.

28. The method of claim 27, wherein enhanced recombination is conferred by increased genomic copy number of the modified or further modified cells.

29. The method of claim 22, wherein the at least one gene is selected from a replication or cell septation gene.

30. The method of claim 29, wherein the modified or further modified cells having evolved toward acquisition of the desired function are selected by their capacity for syncytium formation or cell fusion.

Plant Cells

31. The method of claim 1, wherein the plurality of cells are plant cells and the desired property is improved resistance to a chemical or microbe, and in the screening the steps, the modified or further modified cells are exposed to the chemical or microbe and modified or further modified cells having evolved toward the acquisition of the desired function are selected by their capacity to survive the exposure.

32. The method of claim 31, wherein the microorganism is a virus, bacterium, fungus or insect.

33. The method of claim 32, wherein the chemical is a viricide, fungicide, insecticide, bactericide or herbicide.

34. The method of claim 33, wherein the chemical is BT-toxin.

35. The method of claim 33, wherein the chemical is glyphosate or atrazine.

36. The method of claim 33, further comprising propagating a plant cell having acquired the desired function

to produce a transgenic plant.

Transgenic animal cell

37. The method of claim 1, wherein the plurality of cells are embryonic cells of an animal, and the method further comprises propagating the transformed cells to transgenic animals.

38. The method of claim 37, wherein the modified cells are screened as components of the transgenic animals.

39. The method of claim 38, further comprising obtaining embryonic cells from the transgenic animals having modified cells evolved toward acquisition of the property and transforming the cells with the further library.

40. The method of claim 37, further comprising isolating DNA from transgenic animals that have evolved toward acquisition of the property and introducing the DNA into fresh embryonic cells.

41. The method of claim 37, wherein the animal is a fish.

42. The method of claim 37, wherein at least one of the fragments encodes a growth hormone and the desired property is increased size of the animal.

43. A method of enhancing tissue-specific expression of a protein in a transgenic animal, comprising:

(1) recombining at least first and second forms of a gene encoding a protein, the forms differing from each other in at least two nucleotides, to produce a library of chimeric genes;

(2) screening the library to identify at least one chimeric gene, which as a component of a transgene, confers enhanced expression of the protein in cells from the tissue relative to a transgene containing the wildtype form of the gene;

(3) recombining the at least one chimeric gene with a further form of the gene, the same or different from the first and second forms, to produce a further library of chimeric genes;

(4) screening the further library for at least one further chimeric gene that as a component of a transgene confers enhanced expression of the protein in cells from the tissue relative to a transgene comprising the chimeric gene in the previous screening step;

(5) repeating (3) and (4), as necessary, until the further chimeric gene confers a desired level of expression in cells from the tissue.

44. The method of claim 43, wherein the at least two forms of a gene differ from each other within a coding sequence.

45. The method of claim 43, wherein the at least two forms of a gene differ from each other within a regulatory sequence.

46. The method of claim 43, wherein the cells are mammary gland cells.

47. The method of claim 43, wherein the transgene comprises a milk-protein enhancer, a milk-protein promoter, a signal sequence and a protein coding sequence in operable linkage.

48. The method of claim 43, whereby the protein and marker are expressed as a fusion protein.

49. The method of claim 48, whereby enhanced expression is determined by detecting the presence of the marker as a component of the fusion protein outside the cell expressing the fusion protein.

Use of Recombinogenic Cells

50. A method of performing *in vivo* recombination, comprising

providing a cell incapable of expressing a cell septation gene;

introducing at least first and second segments from at least one gene into a cell, the segments differing from each other in at least two nucleotides, whereby the segments recombine to produce a library of chimeric genes;

selecting a chimeric gene from the library having an

acquired function.

51. The method of claim 50, wherein the cell contains a construct expressing antisense mRNA of the cell septation gene preventing expression of the septation gene.

52. The method of claim 50, wherein the cell is exposed to a drug rendering it incapable of expressing the cell septation gene.

53. The method of claim 50, wherein the cell septation gene contains a mutation preventing its expression.

54. A method of predicting efficacy of a drug in treating a viral infection, comprising

(1) recombining a nucleic acid segment from a virus, whose infection is inhibited by a drug, with at least a second nucleic acid segment from the virus, the second nucleic acid segment differing from the nucleic acid segment in at least two nucleotides, to produce a library of recombinant nucleic acid segments;

(2) contacting host cells with a collection of viruses having genomes including the recombinant nucleic acid segments in a media containing the drug, and collecting progeny viruses resulting from infection of the host cells,

(3) recombining a recombinant DNA segment from a first progeny virus with at least a recombinant DNA segment from a second progeny virus to produce a further library of recombinant nucleic acid segments;

(4) contacting host cells with a collection of viruses having genomes including the further library or recombinant nucleic acid segments, in media containing the drug, and collecting further progeny viruses produced by the host cells,

(5) repeating (3) and (4), as necessary, until a further progeny virus has acquired a desired degree of resistance to the drug, whereby the degree of resistance acquired and the number of repetitions of (3) and (4) needed

to acquire it provide a measure of the efficacy of the drug in treating the virus.

55. The method of claim 54, wherein the media contains a combination of drugs.

56. The method of claim 55, wherein the virus is HIV.

57. A method of predicting efficacy of a drug in treating an infection by a pathogenic microorganism, comprising

(1) transforming a plurality of cells of the microorganism with a library of DNA fragments at least some of which undergo recombination with segments in the genome of the cells to produce modified microorganism cells;

(2) propagating modified microorganisms in a media containing the drug, and recovering surviving microorganisms;

(3) recombining DNA from surviving microorganisms with a further library of DNA fragments at least some of which undergo recombination with cognate segments in the DNA from the surviving microorganisms to produce further modified microorganisms cells;

(4) propagating further modified microorganisms, in media containing the drug, and collecting further surviving microorganisms;

(5) repeating (3) and (4), as necessary, until a further surviving microorganism has acquired a desired degree of resistance to the drug, whereby the degree of resistance acquired and the number of repetitions of (3) and (4) needed to acquire it provide a measure of the efficacy of the drug in killing the pathogenic microorganism.

58. The method of claim 57, further comprising dividing surviving microorganisms into first and second pools, isolating the further library of DNA from the first pool and transforming the second pool with the further library.

59. The method of claim 57, wherein the further library of DNA is obtained from a different microorganism.

Inducing Genetic Exchange

60. A method of evolving a cell to acquire a desired function, comprising:

(a) providing a populating of different cells;

(b) culturing the cells under conditions whereby DNA is exchanged between cells, forming cells with hybrid genomes;

(c) screening or selecting the cells for cells that have evolved toward acquisition of a desired property;

(d) repeating steps (b) and (c) with the selected or screened cells forming the population of different cells until a cell has acquired the desired property.

61. The method of claim 60, wherein DNA is exchanged between the cells by conjugation.

62. The method of claim 60, wherein DNA is exchanged between the cells by phage-mediated transduction.

63. The method of claim 60, wherein DNA is exchnaged between the cells by fusion of protoplasts of the cells.

64. The method of claim 60, wherein DNA is exchanged between the cells by sexual recombination of the cells.

65. The method of claim 60, further comprising transforming a DNA library into the cells.

Protoplast fusion

66. A method of evolving a cell to acquire a desired property, comprising:

(1) forming protoplasts of a population of different cells;

(2) fusing the protoplasts to form hybrid protoplasts, in which genomes from the protoplasts recombine to form hybrid genomes;

(3) incubating the hybrid protoplasts under conditions promoting regeneration of cells;

(4) selecting or screening to isolate regenerated cells that have evolved toward acquisition of the desired property;

(5) repeating steps (1)-(4) with regenerated cells in step (4) being used to form the protoplasts in step (1) until the regenerated cells have acquired the desired property.

67. The method of claim 66, wherein the different cells are fungi cells, and the regenerated cells are fungi mycelia.

68. The method of claim 66, further comprising selecting or screening to isolate regenerated cells with hybrid genomes free from cells with parental genomes.

69. The method of claim 66, wherein a first subpopulation of cells contain a first marker and the second subpopulation of cells contain a second marker, and the method further comprising selecting or screening to identify regenerated cells expressing both the first and second marker.

70. The method of claim 66, wherein the first marker is a membrane marker and the second marker is a genetic marker.

71. The method of claim 69, wherein the first marker is a first subunit of a heteromeric enzyme and the second marker is a second subunit of the heteromeric enzyme.

72. The method of claim 66 further comprising transforming protoplasts with a library of DNA fragments in at least one cycle.

73. The method of claim 72, wherein the DNA fragments are accompanied by a restriction enzyme.

74. The method of claim 66, further comprising exposing the protoplasts to ultraviolet irradiation in at least one cycle.

75. The method of claim 67, wherein protoplasts are provided by treating mycelia or spores with an enzyme.

76. The method of claim 67, wherein the fungus is a fragile strain, lacking capacity for intact cell wall synthesis, whereby protoplast form spontaneously.

77. The method of claim 67, further comprising treating the mycelia with an inhibitor of cell wall formation to generate protoplasts.

78. The method of claim 66, wherein the desired property is the expression of a protein or secondary metabolite.

79. The method of claim 66, wherein the desired property is the secretion of a protein or secondary metabolite.

80. The method of claim 79, wherein the secondary metabolite is taxol.

81. The method of claim 66, wherein the desired property is capacity for meiosis.

82. The method of claim 66, wherein the desired property is compatibility to form a heterokaryon with another strain.

83. The method of claim 67 further comprising exposing the protoplasts or mycelia to a mutagenic agent in at

least one cycle.

Liposome-protoplast fusion

84. A method of evolving a cell toward acquisition of a desired property comprising:

(a) providing a population of different cells;

(b) isolating DNA from a first subpopulation of the different cells and encapsulating the DNA in liposomes;

(c) forming protoplasts from a second subpopulation of the different cells;

(d) fusing the liposomes with the protoplasts whereby DNA from the liposomes is taken up by the protoplasts and recombines with the genomes of the protoplasts;

(e) incubating the protoplasts under regenerating conditions;

(f) selecting or screeing for regenerating or regenerated cells that have evolved toward the desired property;

(g) repeating steps (a)-(f) with the cells that have evolved toward the desired property forming the population of different cells in step (a).

Artificial chromosomes

85. A method of evolving a cell toward acquisition of a desired property, comprising:

(a) introducing a DNA fragment library cloned into an artificial chromosome into a population of cells;

(b) culturing the cells under conditions whereby sexual recombination occurs between the cells, whereby DNA fragments cloned into the artificial chromosome homologously recombine with corresponding segments of endogenous chromosomes of the populations of cells, and endogenous chromosomes recombine with each other;

(c) screening or selecting for cells that have evolved toward acquisition of the desired property.

86. The method of claim 89, wherein the cells are yeast cells and the artificial chromosome is a YAC.

87. The method claim 89, further comprising:

(d) culturing the cells surviving the screening or selecting step under conditions whereby sexual recombination occurs between cells, whereby further recombination occurs between endogenous chromosomes;

(e) screening or selecting for further cells that have evolved toward acquition of the desired property;

(f) repeating steps (d) and (e) as needed until the desired property has been acquired.

88. A method of evolving a DNA segment for acquisition of a desired property, comprising:

(a) providing a library of variants of the segment, each variant cloned into separate copies of an artificial chromosome;

(b) introducing the copies of the artificial chromosome into a population of cells;

(c) culturing the cells under conditions whereby sexual recombination occurs between cells and homologous recombination occurs between copies of the artificial chromosome bearing the variants;

(d) screening or selecting for variants that have evolved toward acquisition of the desired property.

89. A recA protein selected from the group consisting of clone 2, clone 4, clone 5, clone 6 and clone 13 shown in Fig. 13.

90. A method of evolving a recA protein to increase recombinogenic activity, comprising:

shuffling a population of nucleic acid segments encoding variants of recA including a nucleic acid segment selected from the group consisting of clone 2, clone 4, clone 5, clone 6 and clone 13 shown in Fig. 12, to produced recombinant segments;

screening or selecting a recombinant segment with increased recombinogenic activity relative to the nucleic acid segment selected from the group.

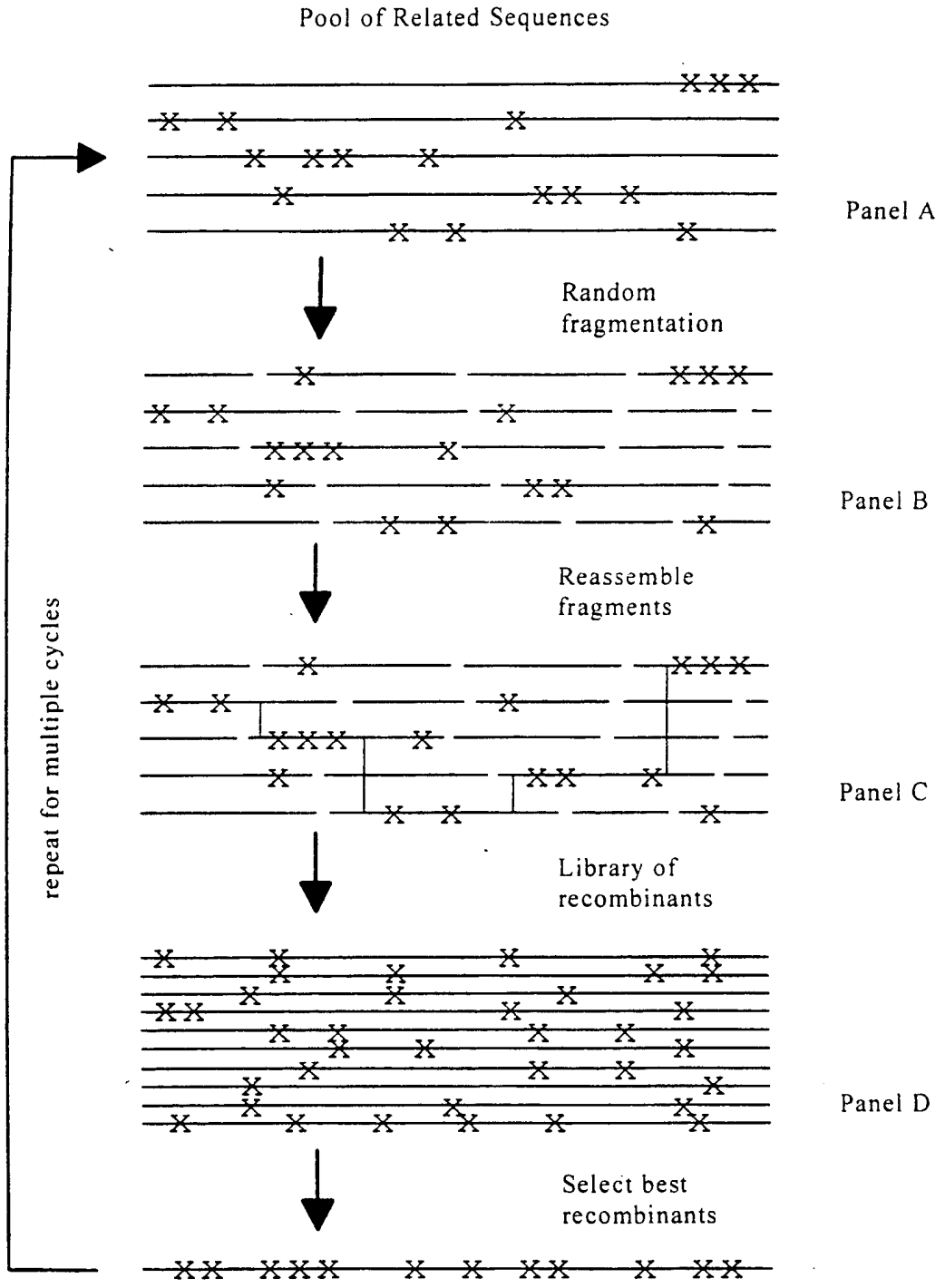

Pool of Related Sequences
Panel A
Random fragmentation
Panel B
Reassemble fragments
Panel C
Library of recombinants
Panel D
Select best recombinants
repeat for multiple cycles

Fig. 1

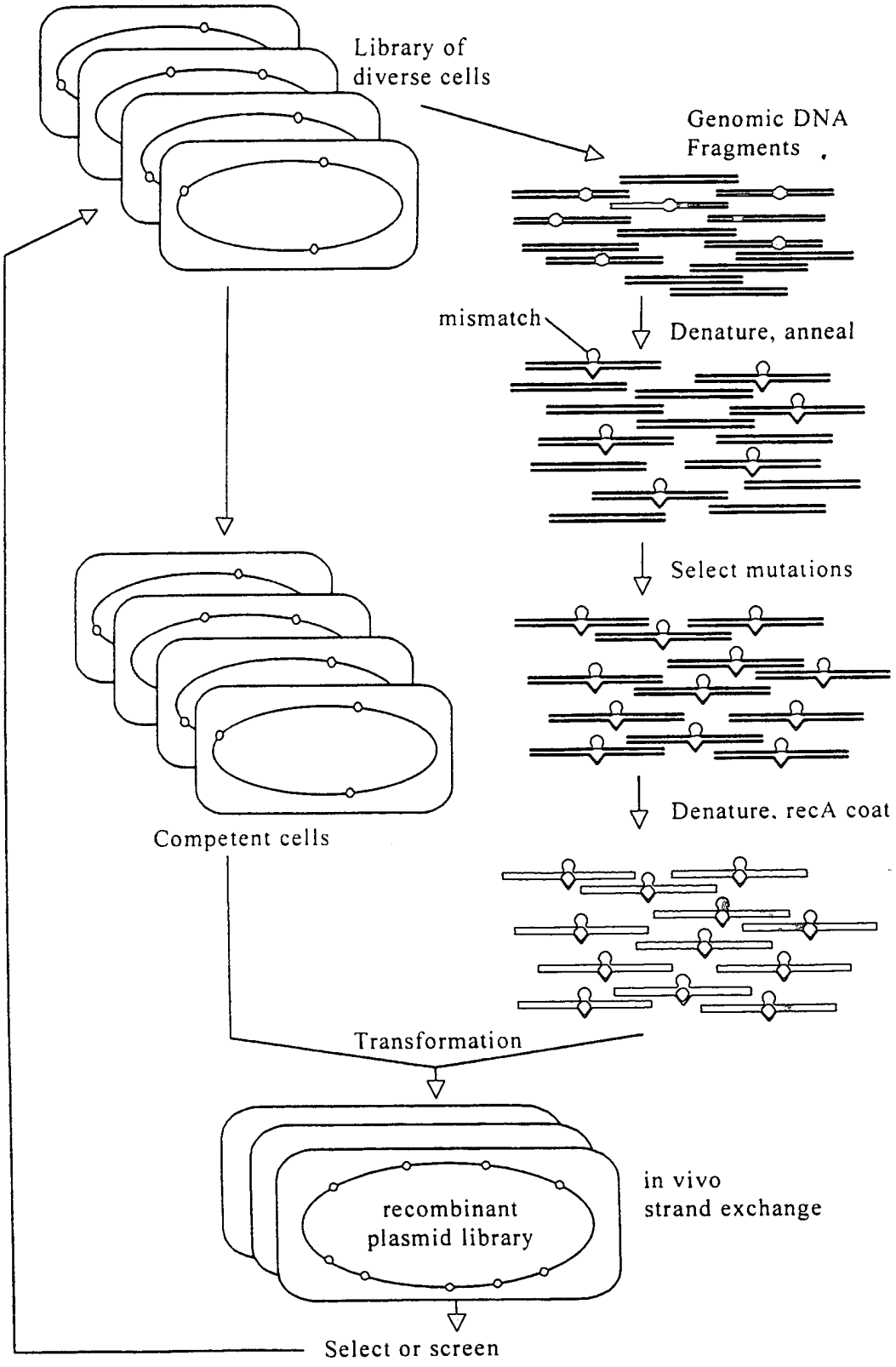

Library of
diverse cells
Genomic DNA
Fragments
mismatch
Denature, anneal
Select mutations
Denature, recA coat
Competent cells
Transformation
recombinant
plasmid library
in vivo
strand exchange
Select or screen

Fig. 2

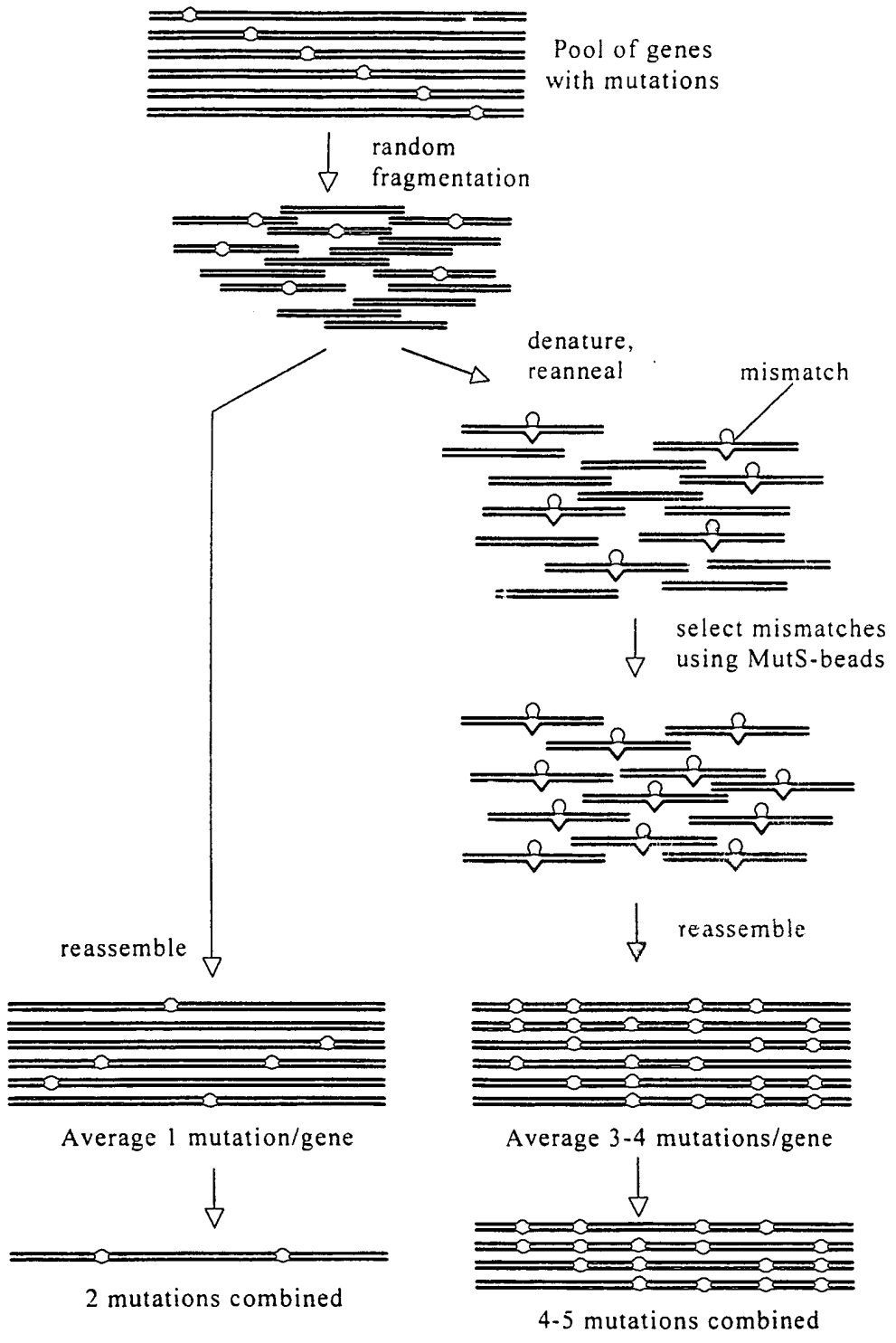

Pool of genes
with mutations
random
fragmentation
denature,
reanneal
mismatch
select mismatches
using MutS-beads
reassemble
reassemble
Average 1 mutation/gene
Average 3-4 mutations/gene
2 mutations combined
4-5 mutations combined

Fig. 3

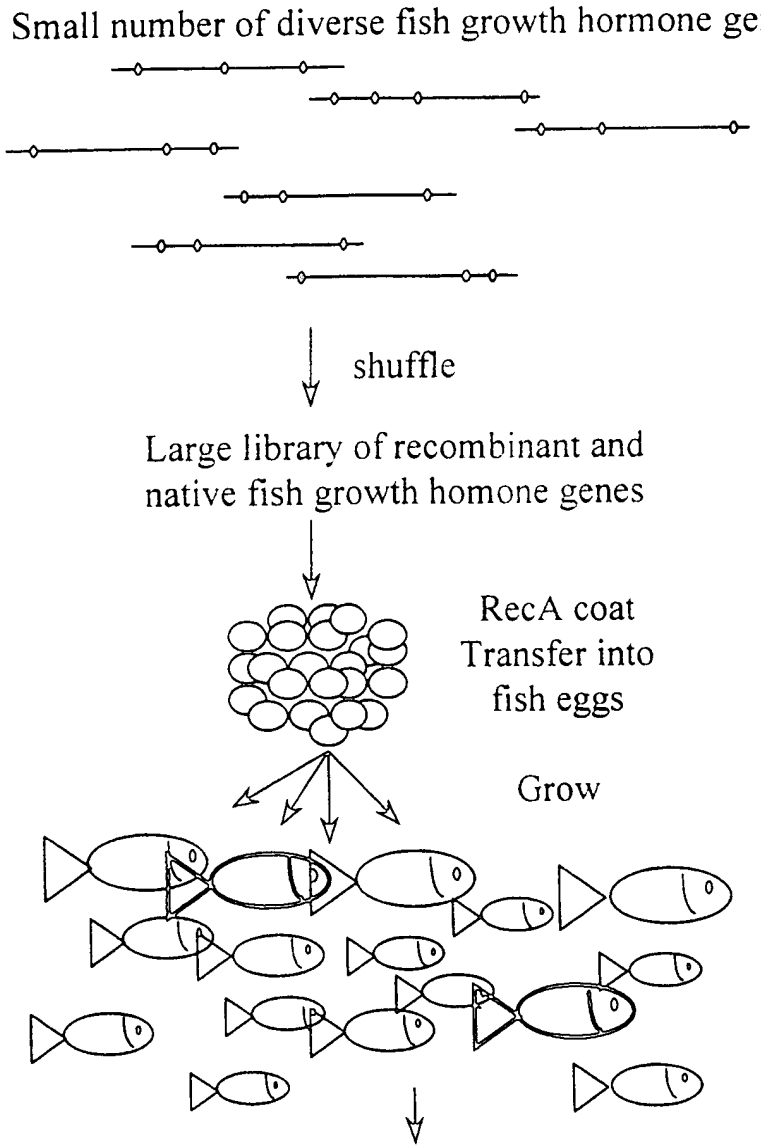

Small number of diverse fish growth hormone genes
shuffle
Large library of recombinant and native fish growth homone genes
RecA coat
Transfer into fish eggs
Grow
select largest fish
PCR out gene

Fig. 4

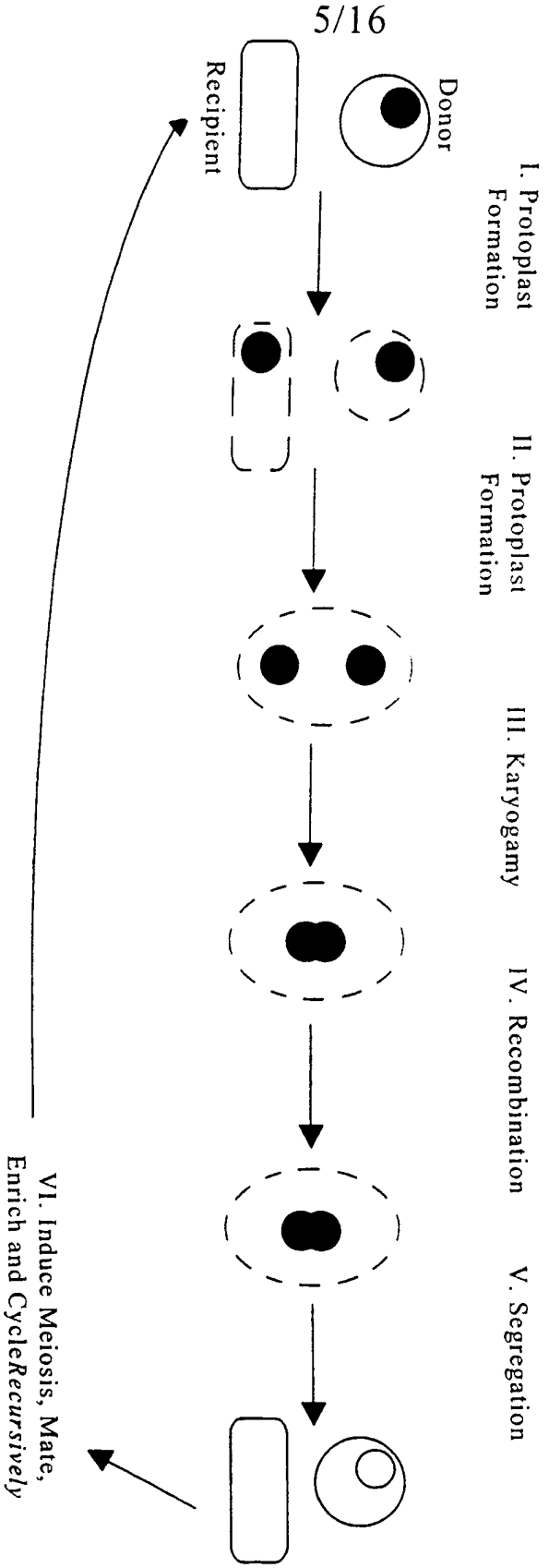

Recipient
Donor
I. Protoplast Formation
II. Protoplast Formation
III. Karyogamy
IV. Recombination
V. Segregation
VI. Induce Meiosis, Mate, Enrich and Cycle*Recursively*

Fig. 5

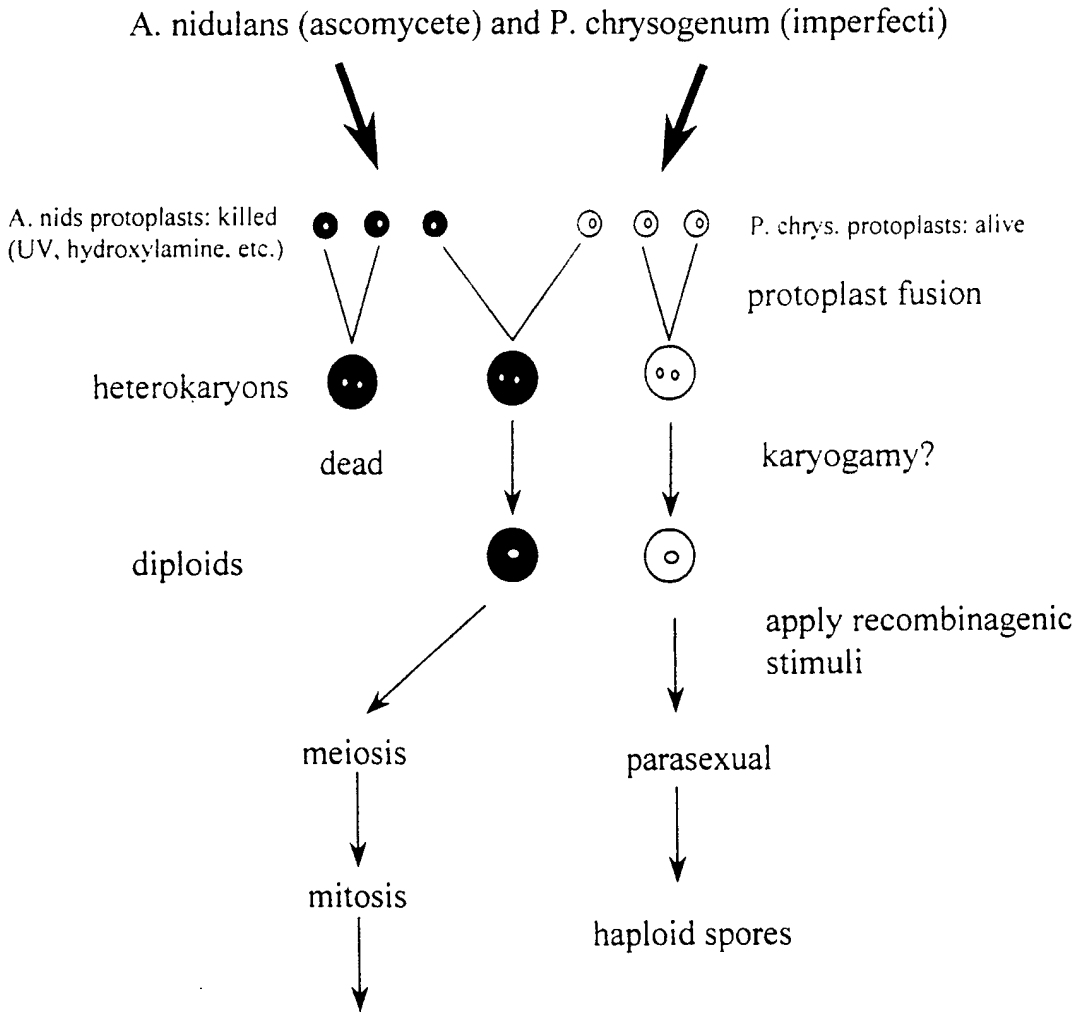

A. nidulans (ascomycete) and P. chrysogenum (imperfecti)
A. nids protoplasts: killed
(UV, hydroxylamine, etc.)
P. chrys. protoplasts: alive
protoplast fusion
heterokaryons
dead
karyogamy?
diploids
apply recombinagenic
stimuli
meiosis
parasexual
mitosis
haploid spores

Fig. 6

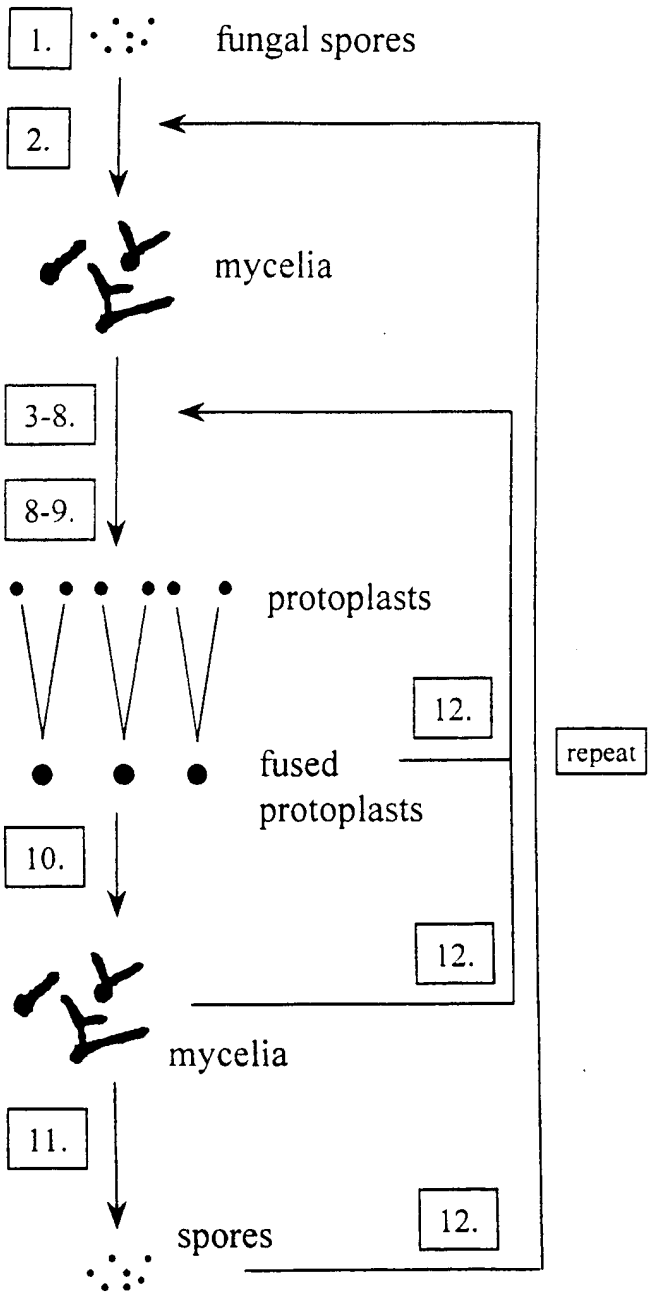
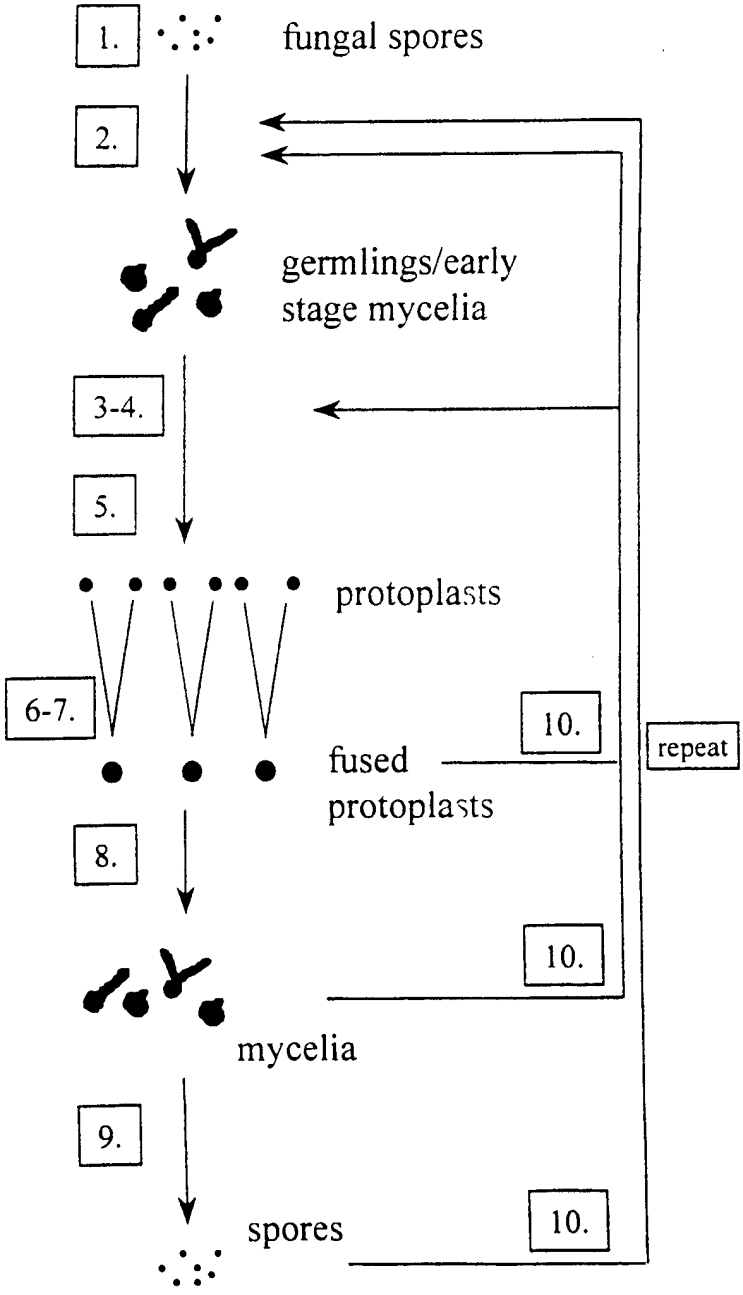

1.
fungal spores
2.
mycelia
3-8.
8-9.
protoplasts
12.
fused
protoplasts
repeat
10.
12.
mycelia
11.
spores
12.

Fig. 7

Fig. 8


1.
fungal spores
2.
germlings/early
stage mycelia
3-4.
5.
protoplasts
6-7.
10.
repeat
fused
protoplasts
8.
10.
mycelia
9.
spores
10.

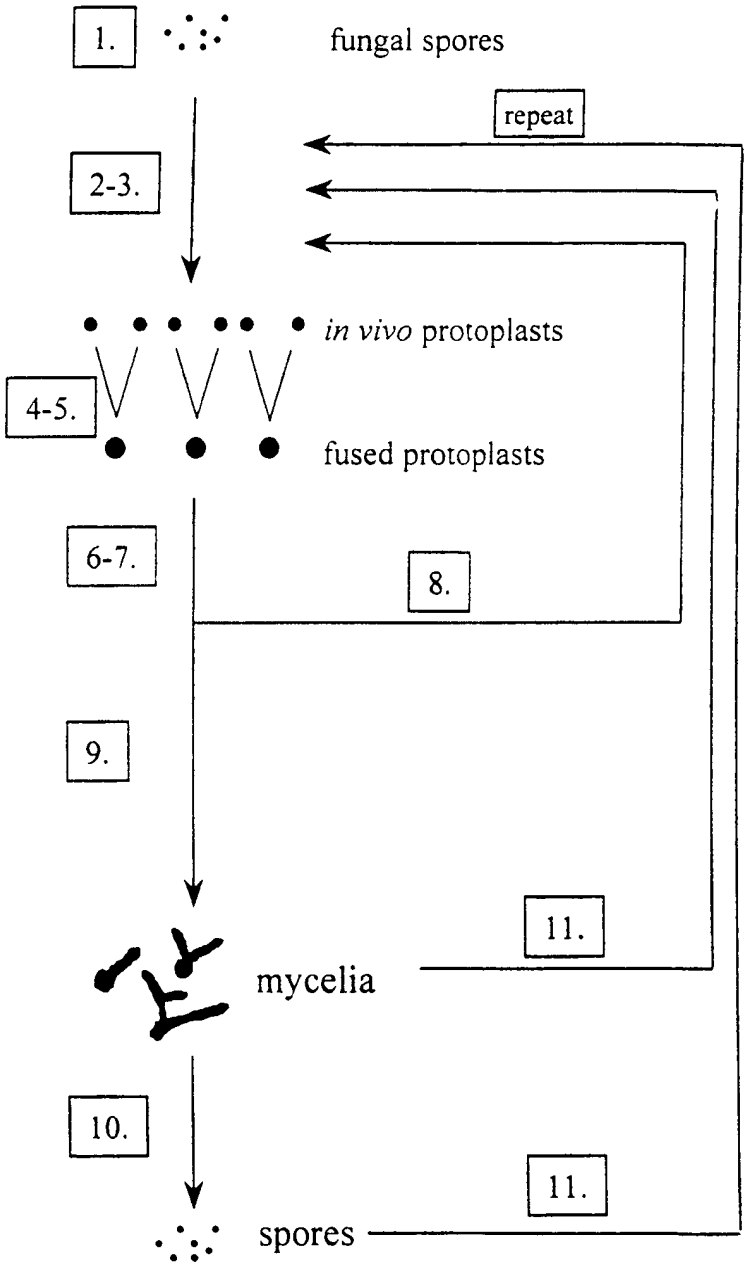

1.
fungal spores
repeat
2-3.
in vivo protoplasts
4-5.
fused protoplasts
6-7.
8.
9.
11.
mycelia
10.
11.
spores

Fig. 9

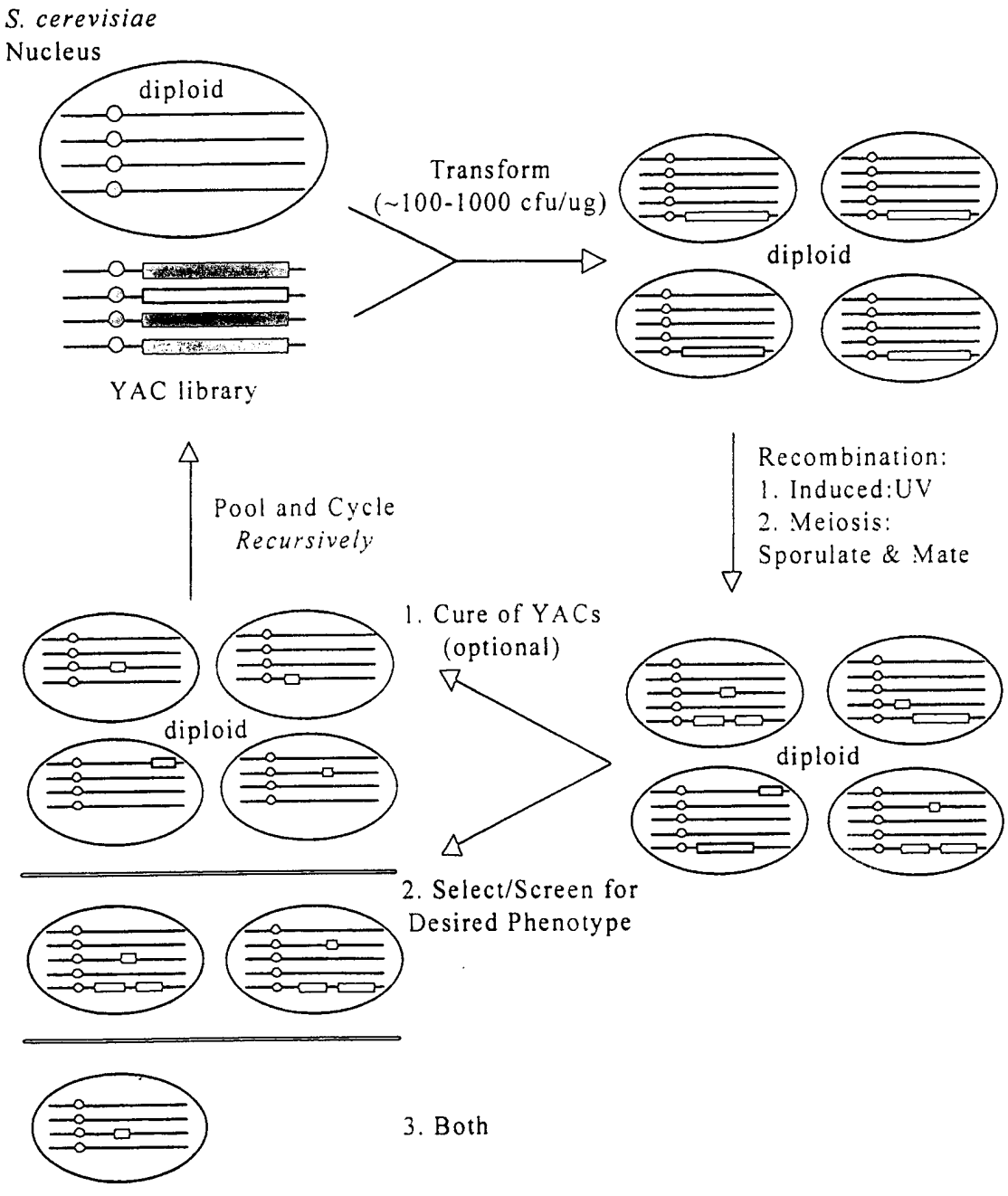

S. cerevisiae
Nucleus
diploid
YAC library
Transform
(~100-1000 cfu/ug)
diploid
Recombination:
1. Induced:UV
2. Meiosis:
Sporulate & Mate
Pool and Cycle
Recursively
1. Cure of YACs
(optional)
diploid
diploid
2. Select/Screen for
Desired Phenotype
3. Both
"regeneration of starting point"

Fig. 10

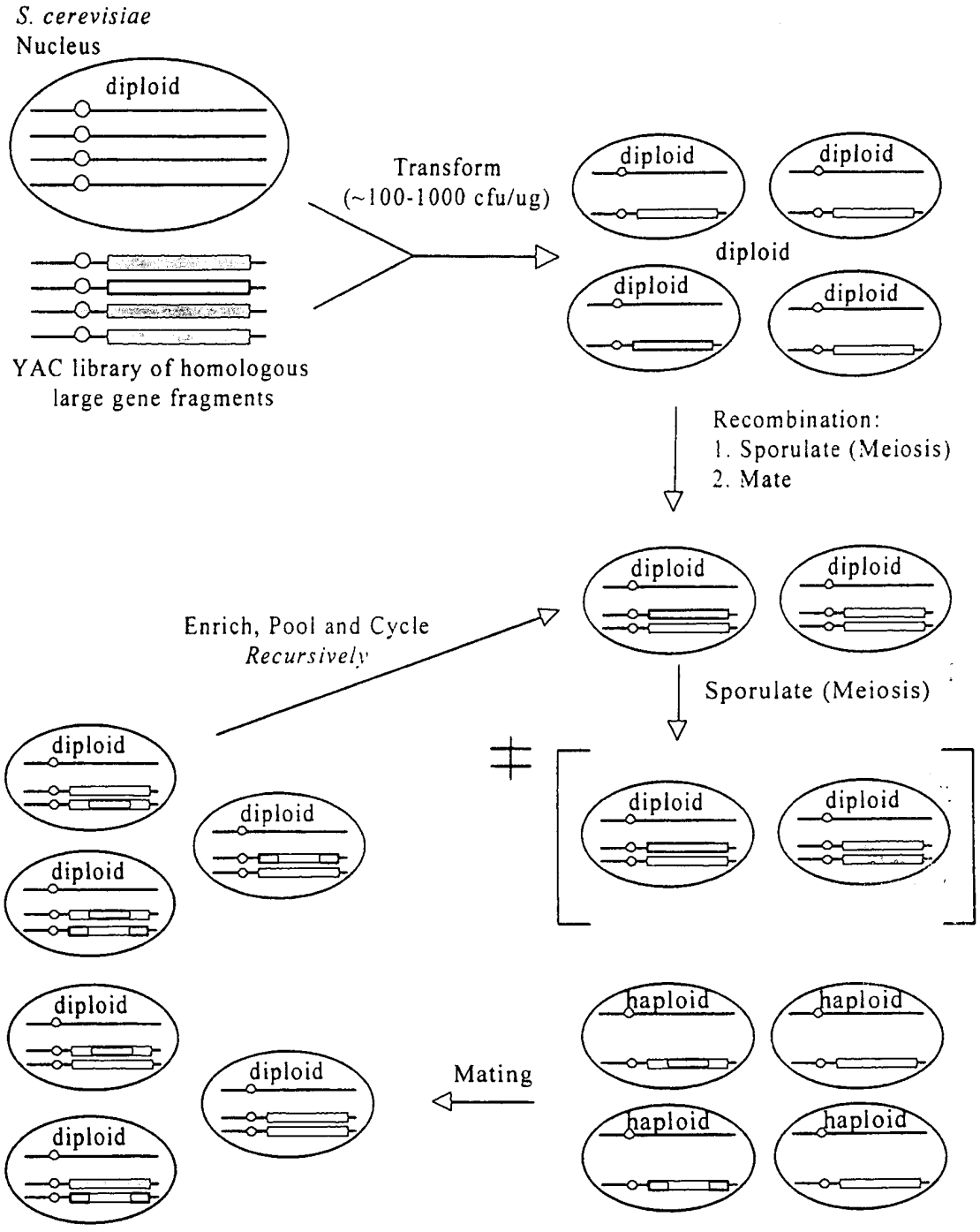

S. cerevisiae
Nucleus
diploid
YAC library of homologous
large gene fragments
Transform
(~100-1000 cfu/ug)
diploid
diploid
diploid
diploid
diploid
Recombination:
1. Sporulate (Meiosis)
2. Mate
diploid
diploid
Enrich, Pool and Cycle
Recursively
Sporulate (Meiosis)
diploid
diploid
diploid
diploid
diploid
diploid
diploid
diploid
haploid
haploid
haploid
haploid
Mating

Fig. 11

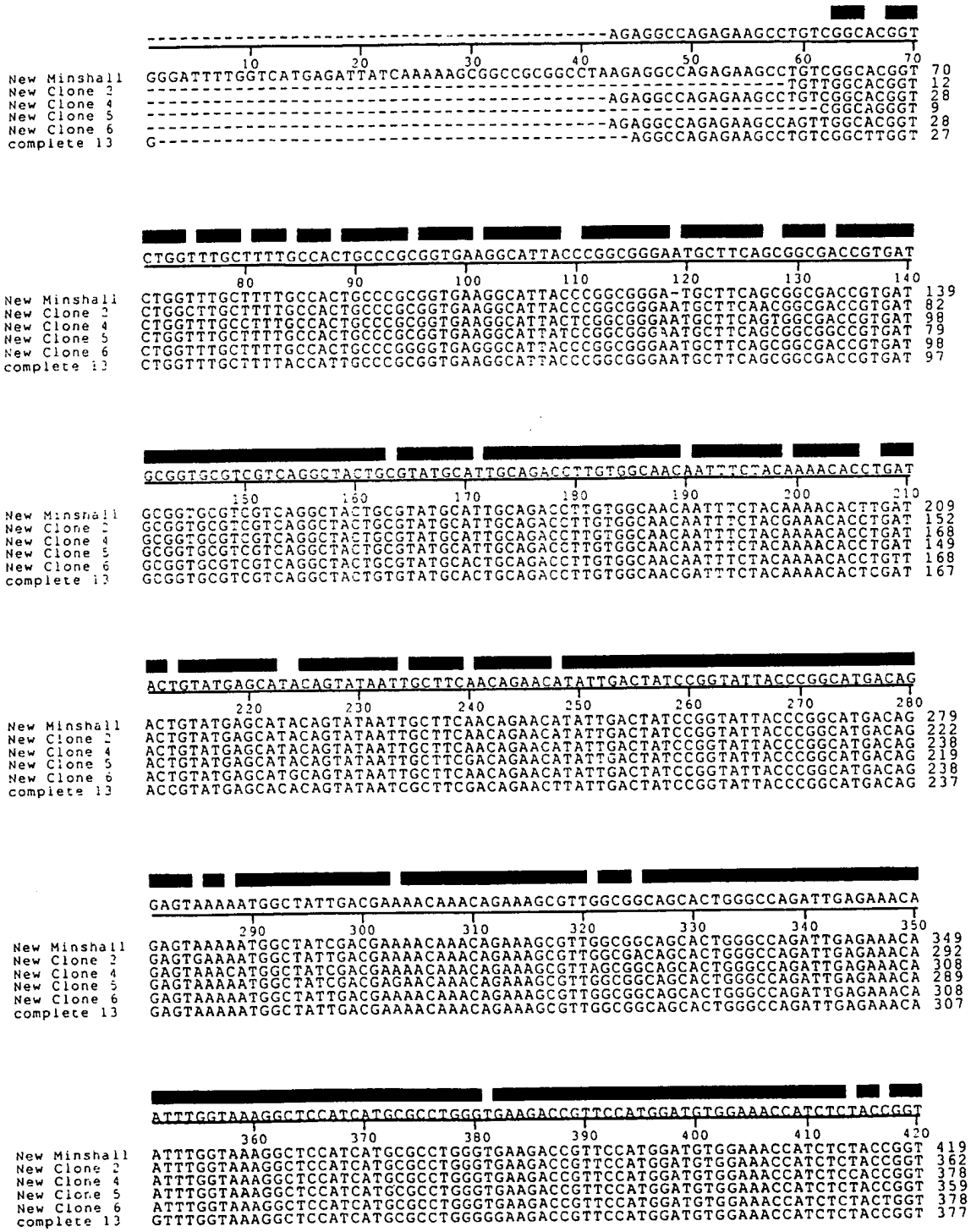


FIG. 12A

SUBSTITUTE SHEET (rule 26)

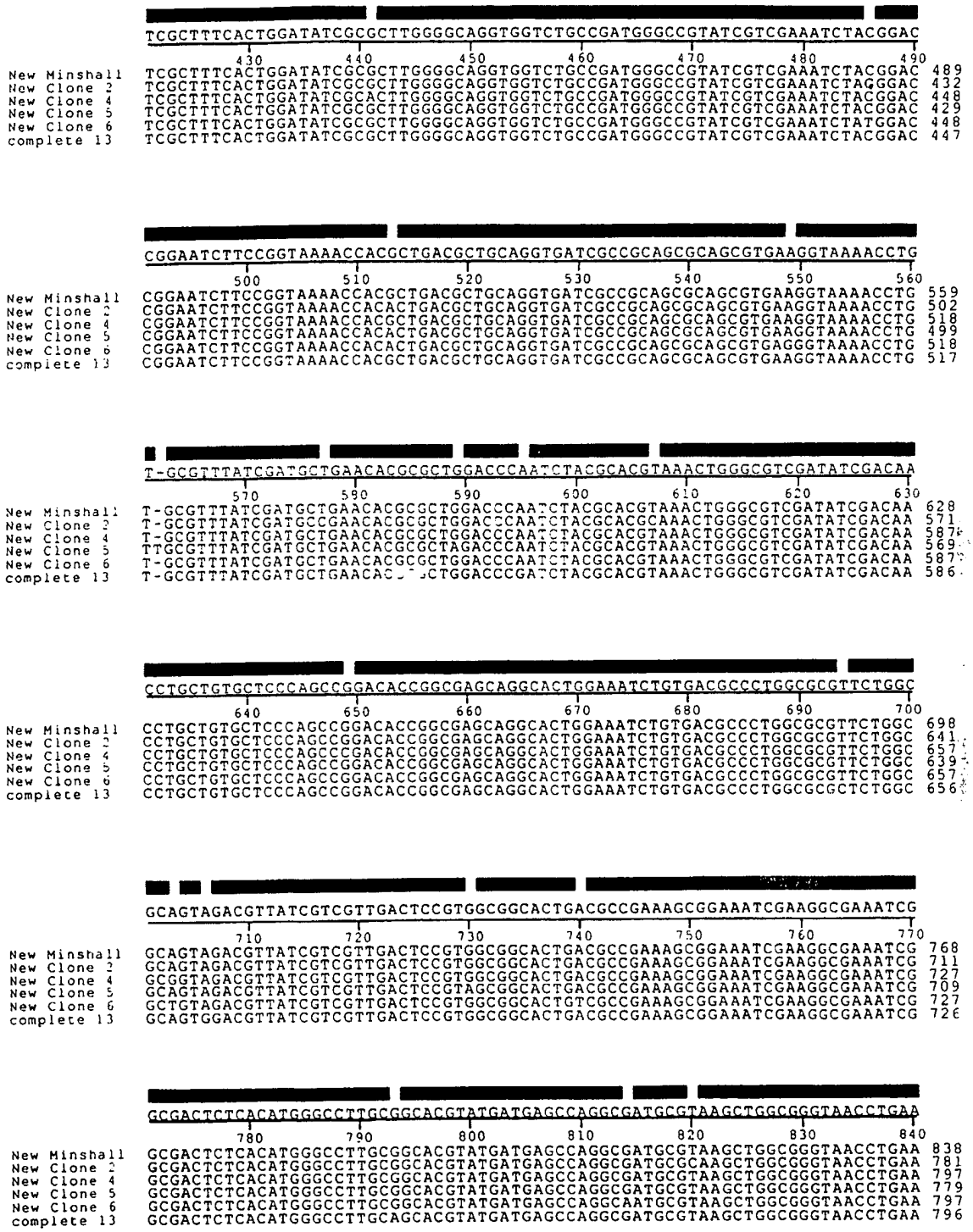


FIG. 12B

SUBSTITUTE SHEET (rule 26)

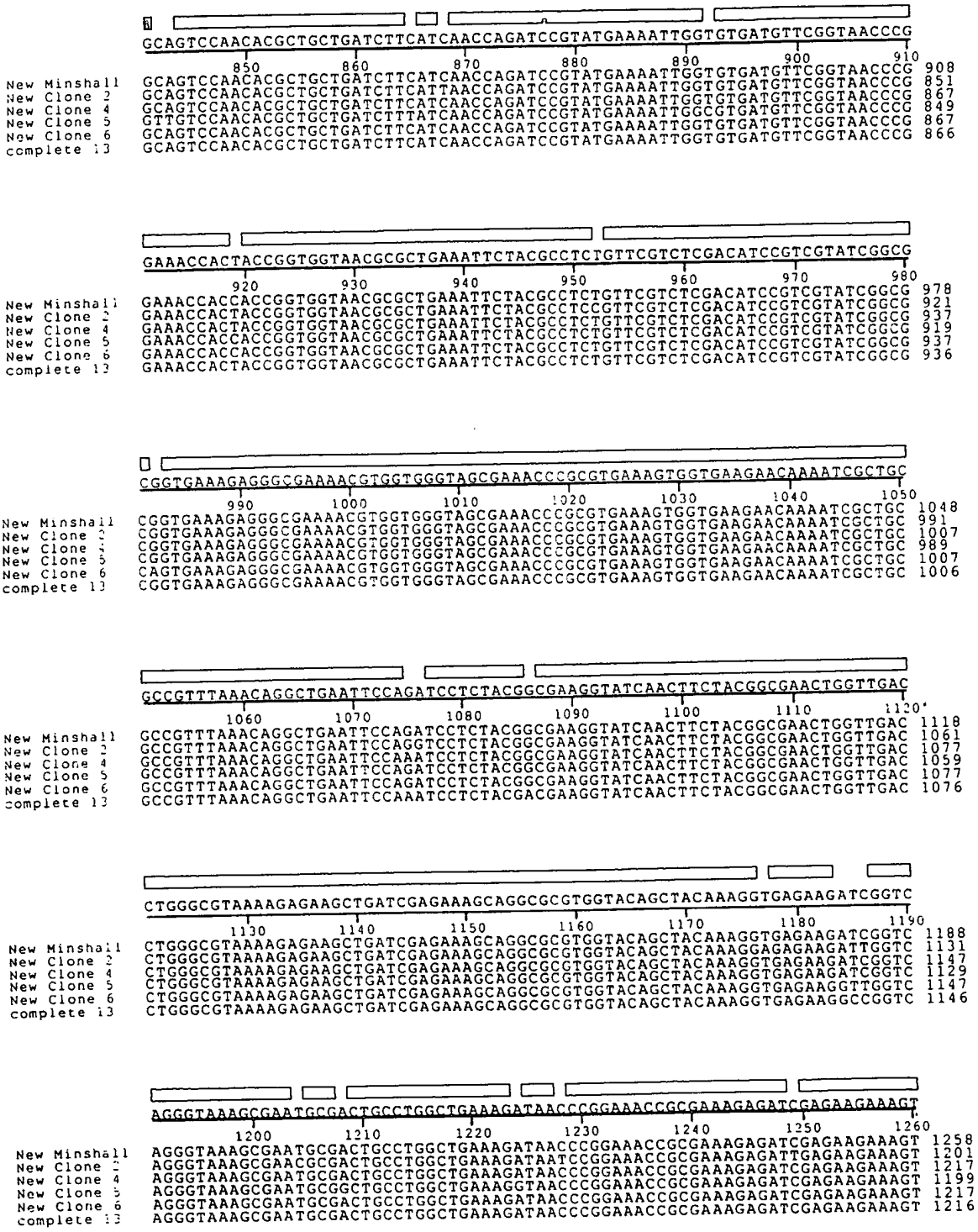


FIG. 12C

SUBSTITUTE SHEET (rule 26)

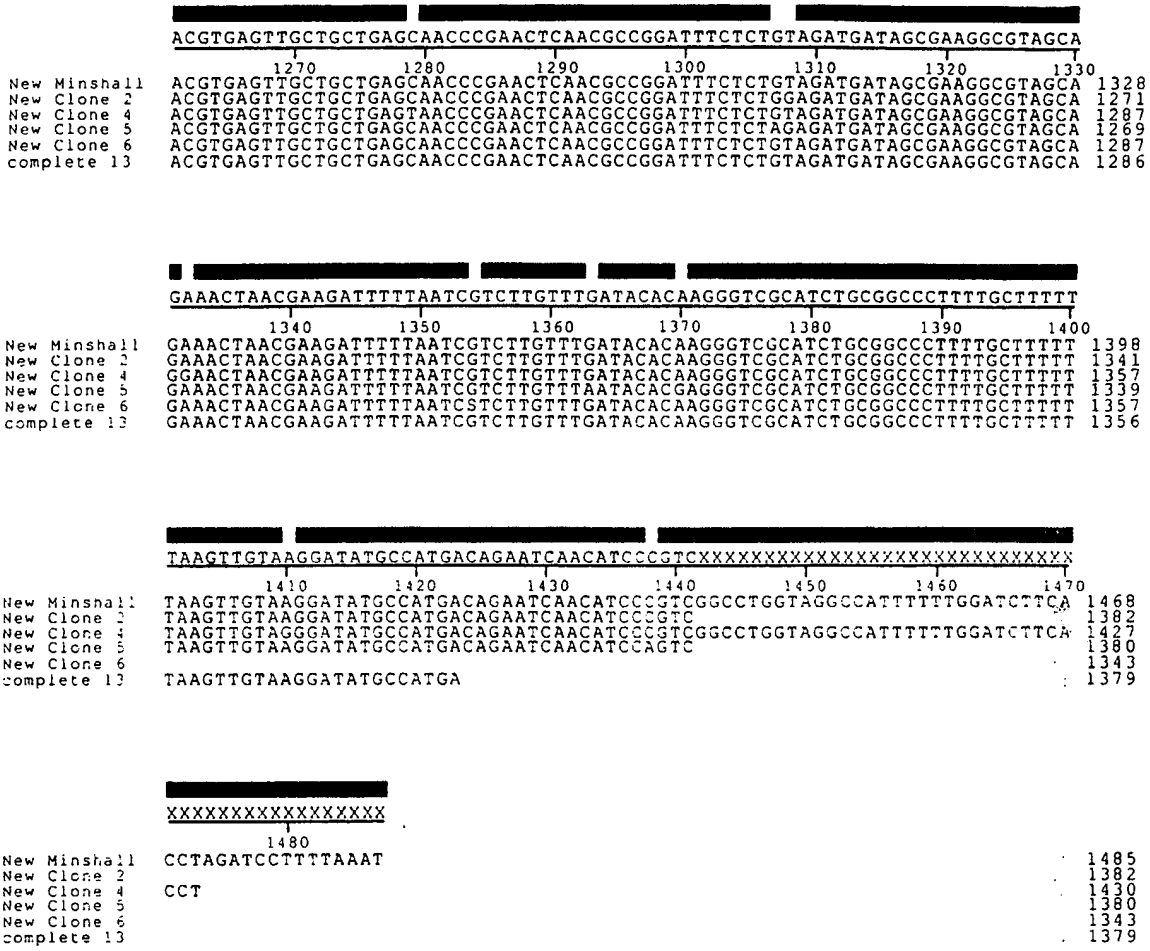


FIG. 12D

SUBSTITUTE SHEET (rule 26)

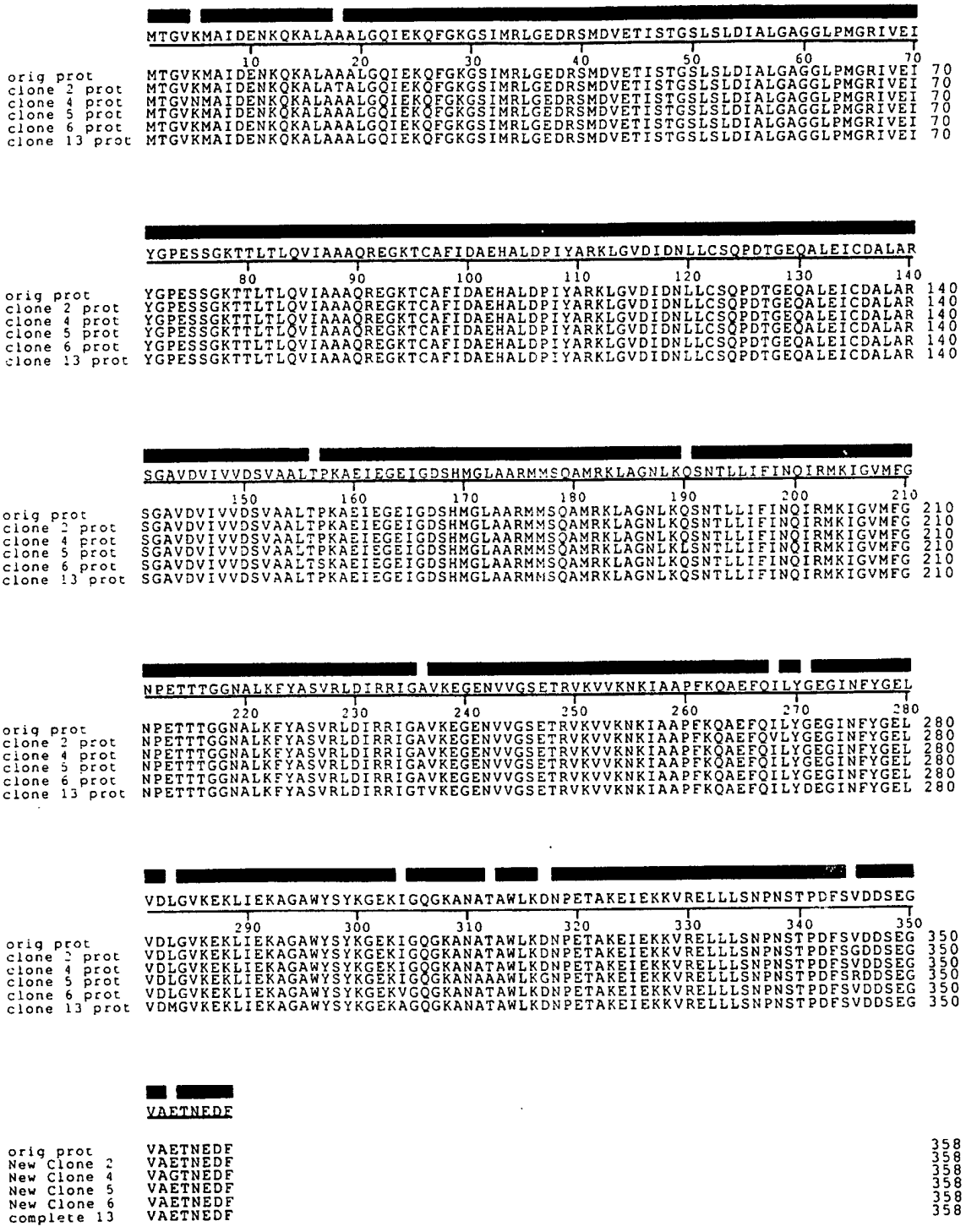

MTGVKMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI
orig prot MTGVKMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
clone 2 prot MTGVKMAIDENKQKALATALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
clone 4 prot MTGVNMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
clone 5 prot MTGVKMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
clone 6 prot MTGVKMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
clone 13 prot MTGVKMAIDENKQKALAAALGQIEKQFGKGSIMRLGEDRSMDVETISTGSLSLDIALGAGGLPMGRIVEI 70
YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR
orig prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
clone 2 prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
clone 4 prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
clone 5 prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
clone 6 prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
clone 13 prot YGPESSGKTTLTLQVIAAAQREGKTCAFIDAEHALDPIYARKLGVDIDNLLCSQPDTGEQALEICDALAR 140
SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG
orig prot SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG 210
clone 2 prot SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG 210
clone 4 prot SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG 210
clone 5 prot SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKLSNTLLIFINQIRMKIGVMFG 210
clone 6 prot SGAVDVIVVDSVAALTSKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG 210
clone 13 prot SGAVDVIVVDSVAALTPKAEIEGEIGDSHMGLAARMMSQAMRKLAGNLKQSNTLLIFINQIRMKIGVMFG 210
NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEGINFYGEL
orig prot NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEGINFYGEL 280
clone 2 prot NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQVLYGEGINFYGEL 280
clone 4 prot NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEGINFYGEL 280
clone 5 prot NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEGINFYGEL 280
clone 6 prot NPETTTGGNALKFYASVRLDIRRIGAVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYGEGINFYGEL 280
clone 13 prot NPETTTGGNALKFYASVRLDIRRIGTVKEGENVVGSETRVKVVKNKIAAPFKQAEFQILYDEGINFYGEL 280
VDLGVKEKLIEKAGAWYSYKGEKIGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEG
orig prot VDLGVKEKLIEKAGAWYSYKGEKIGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEG 350
clone 2 prot VDLGVKEKLIEKAGAWYSYKGEKIGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSGDDSEG 350
clone 4 prot VDLGVKEKLIEKAGAWYSYKGEKIGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEG 350
clone 5 prot VDLGVKEKLIEKAGAWYSYKGEKIGQGKANAAAWLKGNPETAKEIEKKVRELLLSNPNSTPDFSRDDSEG 350
clone 6 prot VDLGVKEKLIEKAGAWYSYKGEKVGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEG 350
clone 13 prot VDMGVKEKLIEKAGAWYSYKGEKAGQGKANATAWLKDNPETAKEIEKKVRELLLSNPNSTPDFSVDDSEG 350
VAETNEDF
orig prot VAETNEDF 358
New Clone 2 VAETNEDF 358
New Clone 4 VAGTNEDF 358
New Clone 5 VAETNEDF 358
New Clone 6 VAETNEDF 358
complete 13 VAETNEDF 358

FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/00852

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; C12N 15/00
US CL : 435/6, 172.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 172.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, STN; MEDLINE, EMBASE, SCISEARCH, CAPLUS, WPIDS, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	STEMMER, W. P. C. DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution. Proc. Natl. Acad. Sci. October 1994, Vol. 91, pages 10747-10751, see entire document.	1-94
Y	STEMMER, W. P. C. Rapid evolution of a protein *in vitro* by DNA shuffling. Nature. 04 August 1994, Vol. 370, pages 389-391, see entire document.	1-94
Y	US 5,521,077 A (KHOSLA et al.) 28 May 1996, see entire document.	1-94

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
03 APRIL 1998	21 MAY 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ANDREW WANG Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*